

A background image showing two healthcare professionals, a man and a woman, in a clinical setting. The man is on the left, wearing a white shirt and a stethoscope, looking towards the woman. The woman is on the right, wearing glasses and a dark top, looking down. The image is overlaid with a blue geometric pattern of triangles.

## Whole population models of provision: Establishing integrated budgets

Accountable Care Organisation (ACO) Contract package  
- supporting document

**Our values:**  
clinical engagement, patient involvement,  
local ownership, national support

August 2017

## Whole population models of provision: Establishing integrated budgets

This document has been written in partnership with NHS Improvement.

**Version number:** 1

**First published:** August 2017

**Gateway publication reference:** 06264

### Equality and health inequalities statement

Promoting equality and addressing health inequalities are at the heart of NHS England's values. Throughout the development of the policies and processes cited in this document, we have:

- given due regard to the need to eliminate discrimination, harassment and victimisation, to advance equality of opportunity, and to foster good relations between people who share a relevant protected characteristic (as cited under the Equality Act 2010) and those who do not share it; and
- given regard to the need to reduce inequalities between patients in access to, and outcomes from healthcare services and to ensure services are provided in an integrated way where this might reduce health inequalities.

# Contents

<b>Contents</b>	<b>3</b>
<b>Summary</b>	<b>5</b>
<b>1. Introduction</b>	<b>7</b>
1.1. Integrated budgets to support implementation of ACO care models	7
1.1.1. Need for a new payment approach	7
1.1.2. Capitation with outcomes	8
1.1.3. Why integrated budgets (WPBs) with outcomes based payment?	9
1.1.4. WPB in the context of the National Tariff and GMS/PMS regulations	9
1.1.5. Essential requirements	10
1.1.6. Commissioner considerations	11
1.2. Overview of WPB development	12
1.2.1. Calculating the WPB baseline	12
1.2.2. Estimating WPB values for future years	13
1.2.3. Converting estimated WPB values into contract values	13
1.2.4. Improvement Payment Scheme	13
1.2.5. Gain/loss sharing	14
<b>2. Calculating the WPB baseline</b>	<b>15</b>
2.1. Introduction	15
2.1.1. Foundational activities	15
2.2. Addressing likely data challenges	15
2.3. Service-level spend calculated from activity-based payments	16
2.4. Service-level spend on services within scope of block contracts	17
2.4.1. Overview of the approach to block contracts	17
2.4.2. Step 1: Allocating block contract (£) amounts to individual services	20
2.4.3. Step 2: Collecting activity counts and mapping them to individual services	25
2.4.4. Step 3: Allocating commissioner spend (£) amounts to individual activity counts	28
2.4.5. Step 4: Mapping the monetised activity counts to individual patients and bundling them into GP lists (see Figure 12)	33
2.5. Service-level spend on primary care	36
2.6. Current spend in social care contracts	37
2.7. Convert the current spend to the WPB baseline	39
<b>3. Estimating WPB values for future years</b>	<b>40</b>
3.1. Introduction	40
3.2. Considerations for forecasting	42
3.2.1. Commissioner allocation growth rates	42
3.2.2. Key factors affecting future WPB values	42
3.3. Step 1: Projecting the WPB baseline	43
3.3.1. Use of commissioner allocation growth rates	43
3.3.2. National planning assumptions	44
3.3.3. Refining national assumptions with local information	49
3.4. Step 2: Adjust the baseline projection to reflect provider efficiencies	51
3.5. Step 3: Adjust to reflect impacts of new care model	51

3.6. Step 4: Assurance and verification	52
3.7. Conclusion	53
<b>4. Converting WPB estimated values into contractual values</b>	<b>54</b>
4.1. Introduction	54
4.2. Limits for agreed contract variations	55
4.3. Periodic agreement of actual contracted WPB values	55
4.3.1. Process for the periodic agreement of the contracted WPB	56
4.4. Mechanistic update for population changes	56
4.5. Conclusion	57
<b>5. Improvement Payment Scheme (IPS)</b>	<b>58</b>
5.1. Introduction	58
5.2. ACO IPS	58
5.3. Scheme quantum	59
5.4. Commissioner flexibilities	59
5.5. IPS timeline	59
<b>6. Gain/loss sharing</b>	<b>60</b>
6.1. Introduction	60
6.2. Definition	60
6.3. Context and purpose	61
6.3.1. Aligning financial incentives	61
6.3.2. Managing the transfer and impact of utilisation risk	62
6.4. Implementation framework	62
6.4.1. Gain/loss sharing implementation framework	64
6.4.2. Step 1: Identifying the activities to include	65
6.4.3. Step 2: Selecting the organisations to include	66
6.4.4. Step 3: Determining the activity profile	67
6.4.5. Step 4: Sharing gains or losses	67
6.4.6. Step 5: Calculating and allocating gains or losses	68
6.4.7. Step 6: Any necessary adjustments	68
6.5. Design parameters	68
6.5.1. Duration of gain/loss sharing arrangement	68
6.5.2. Further detail on how gains/losses will be shared	70
6.6. Next steps	73

## Summary

The vision for health and social care in England presented in the NHS Five Year Forward View<sup>1</sup> highlighted three 'gaps' that need to be addressed:

- funding and efficiency
- health and wellbeing
- care and quality.

The NHS Five Year Forward View promoted new care models, including multispecialty community providers (MCP) and integrated primary and acute care systems (PACS), to close these gaps. Successful implementation of these Accountable Care Organisation (ACO) models requires a new payment approach that will provide financial incentives to facilitate greater co-ordination and integration of care.

This handbook describes the development of a payment approach for whole population models of integrated provision, centred on integrated budgets derived from current commissioner expenditure. These are commonly described in the sector as whole population budgets (WPB) and we therefore use this term for consistency. This does not imply that the budget provides for all of the services delivered to an individual, rather that this is a budget for the whole of population served by the relevant provider, across the services in scope of its contract. This approach has been developed to encourage the promotion of whole population management, prevention, self-care and a focus on outcomes rather than inputs. The WPB approach promotes that the whole population is incorporated into the budget, but allows for flexibility of service scope to account for different care models being developed to meet local needs.

The WPB approach is characterised by the following features:

- covers the relevant service scope for the whole population (ie the registered list of patients and those in the locality of the care model not registered with GPs, rather than a segment of the population)<sup>2</sup>
- removes the direct relationship between activity and payment
- improves alignment of payment for all providers within the care model
- better incentivises prevention and wellbeing
- focuses on management of outcomes, activity and costs across the system.

NHS England and NHS Improvement are working with a number of MCP and PACS vanguards to demonstrate how the constituent parts of a WPB can be developed at pace, recognising that both an MCP and a PACS underpinned by a new contract are types of ACO.

There are three elements to the overall payment approach for ACO care models:

1. Establishing the required WPB for delivering services:
  - calculating the baseline contractual value of services in scope of the ACO care model
  - estimating WPB values for the full contract term to support multiyear contracting
  - converting estimated WPB values to contractual values.

<sup>1</sup> <https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf>

<sup>2</sup> While an ACO may start with a segmented approach, we expect a whole population approach to be adopted as these new care models mature.

2. Designing an Improvement Payment Scheme (IPS) for ACO providers that will operate as a top-slice from the WPB.
3. Developing gain/loss sharing to:
  - build and align financial incentives across local areas
  - manage the transfer of utilisation risk from commissioner to provider that is associated with implementing a WPB.

In the short to medium term,<sup>3</sup> given current system capability and data availability, we consider this approach to be the most effective and practical way of facilitating ACO implementation.

We have developed this handbook to support local development of a WPB using what we have learned from working with vanguards. We include case studies to illustrate the main stages in developing a WPB for an ACO, from design considerations through to WPB calculation, and tools and templates to help you on this path.

The handbook should be used alongside the ACO emerging care model and contract framework document, as well as the emerging contract itself.<sup>4,5</sup>

### **Contract value and possible variations to value and scope during the contract term**

From a procurement perspective, to be transparent, it is important that the procurement documents for the relevant contract set out the indicative annual contract values and the way in which payment may be made (including any possible extensions/variations).

Commissioners may, for example, wish to have dialogue with the bidders in relation to payment mechanisms as part of the procurement process. Commissioners may do so taking account of their obligations to be transparent, treat bidders equally and in a non-discriminatory manner.

To minimise any potential procurement risk that contract value or payment changes during the term of the contract could trigger, then in their procurement documents, commissioners should build in the fact that the contract value or payment mechanism may be subject to change, and provide a list of non-exhaustive options as to the types of changes that may occur during the term of the contract. Please refer to the *Procurement and assurance approach*<sup>6</sup> for more information, including the possible variations that are permitted by current procurement legislation.

3 The time between implementing a WPB and transition to a full capitated payment system will depend on the business case for change, and on the availability of accurate patient-level costing information and may vary between organisations.

4 <https://www.england.nhs.uk/wp-content/uploads/2016/07/mcp-care-model-frmwrk.pdf>

5 <https://www.england.nhs.uk/newbusinessmodels/publications>

6 <https://www.england.nhs.uk/publication/procurement-and-assurance-approach-document-4>

# 1. Introduction

The handbook has been jointly produced by NHS England and NHS Improvement. It provides practical guidance to support local areas looking to implement an ACO care model and, specifically, a Whole Population Budget (WPB) approach to paying for health and care services.

This handbook includes information gained through working with vanguards and provides guidance on a number of technically challenging aspects of designing and calculating a WPB, building on previous publications on capitation.<sup>7,8,9</sup>

Each chapter is modular so the reader can access individual chapters without necessarily reading the whole document. Guidance is provided as follows:

- Chapter 2: how to calculate the WPB baseline for an ACO.
- Chapter 3: how to estimate WPB values for the duration of the ACO multiyear contract.
- Chapter 4: how to convert estimated WPB values into contractual values, including agreeing a mechanism to adjust the WPB amount during the contract where necessary to maintain the intended effects.
- Chapter 5: how to implement the IPS.
- Chapter 6: how to implement gain/loss sharing to better align financial incentives and to help to manage the transfer of utilisation risk from commissioner to provider.

## 1.1. Integrated budgets to support implementation of ACO care models

### 1.1.1. Need for a new payment approach

A new payment mechanism is needed to maximise the success of ACO care models in delivering whole population integrated care. Sector feedback indicates that the current payment system does not always support delivery of more integrated and better co-ordinated care centred on the patient.

Activity-based payment models emphasise increased volume of activity rather than incentivising prevention or delivery of better outcomes. Block contracts provide a consolidated fixed payment irrespective of activity, with no incentive to deliver services or activity beyond those specified in the contract. Further, block contracts often do not provide the transparency necessary for continuous improvement.

In the near term we propose WPBs, a version of a capitation-based payment, are implemented as a pragmatic payment approach to facilitate delivery of integrated care models.

<sup>7</sup> <https://www.gov.uk/guidance/capitation>

<sup>8</sup> [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/445731/LPE\\_Capitation.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/445731/LPE_Capitation.pdf)

<sup>9</sup> [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/445741/Capitation\\_payment\\_-\\_international\\_examples.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/445741/Capitation_payment_-_international_examples.pdf)



### 1.1.2. Capitation with outcomes

A capitated outcomes-based payment model is one approach to the payment of integrated health and care services delivered by an ACO provider. NHS England and NHS Improvement have documented the international evidence<sup>10</sup> supporting capitation, and provided local payment examples<sup>11</sup> that illustrate how it may apply in England and how it could be combined with gain/loss sharing to align and manage system-wide incentives.<sup>12</sup>

Capitated approaches should have the following core characteristics:

- **Predictability:**
  - to increase system stability to plan and implement changes. The value of the budget is identified upfront and set in the context of a multiyear contract.
- **Accountability and flexibility:**
  - to increase provider accountability for the holistic care needs of individuals by providing an incentive to co-ordinate care across settings and providers
  - to increase opportunities to change service delivery across care pathways centred on patients
  - potentially to reduce complexity of commissioner and provider relationships.
- **Risk and reward:**
  - to incentivise investment in preventive care and treatment in the appropriate lowest cost setting
  - to reward providers for doing the right thing but without specifying exactly what is done or how care is delivered.

A capitated approach works by enabling risk to be more appropriately allocated to the organisation best placed to influence, manage and bear specific types of risk. The most sophisticated capitation models include risk adjustments to reflect the differing needs of individuals within the population, use patient-level information and consider utilisation of the full range of services covered by the payment.

Capitation aims to allocate risk as follows:

- **Risks borne by commissioners:**
  - Population size risk: The risk that the population within scope of the new care model is materially larger or smaller than initially assumed.
  - Epidemiological and demographic risk: The risk that the composition and health needs of the population materially differ from those initially assumed; for example, it has a higher proportion of elderly patients or a greater prevalence of long-term conditions.
- **Risks borne by providers:**
  - Utilisation risk: The risk that services are utilised more or less frequently than predicted given the population size and profile.
  - Volatility risk: The risk that random variation in usage patterns may result in different levels of service use.
  - Efficiency risk: The risk that the intended efficiency gains are not realised, or that the unit cost of services is not as expected.
  - Quality risk: The risk that a system-wide change adversely affects the quality of care provided, and in turn patient health outcomes.

10 [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/445741/Capitation\\_payment\\_-\\_international\\_examples.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/445741/Capitation_payment_-_international_examples.pdf)

11 [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/445731/LPE\\_Capitation.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/445731/LPE_Capitation.pdf)

12 [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/445617/GLLPE.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/445617/GLLPE.pdf)



### 1.1.3. Why integrated budgets (WPBs) with outcomes based payment?

WPBs are a simplified version of capitated payments, providing periodic payments for a range of services, initially based on current commissioner spend, according to the size and needs of the population.<sup>13</sup> In the near term, WPBs are the pragmatic payment solution for the following reasons:

- The ACO care models cover a whole population. Matching this coverage in payment terms offers greater opportunities to integrate care and incentivise prevention.
- Data and capability constraints, particularly outside of acute care, limit the ability to implement a formula-based or person-level capitated payment mechanism at this time. The whole population approach can more readily use data currently available in the NHS as a starting point for calculation of payment.
- It is easier to operate than a population segment-based approach, as individuals and the care they receive do not need to be allocated to mutually exclusive segments, and movement of individuals between segments does not need to be regularly tracked.

The WPB approach requires more effort to maintain from year to year than person-level capitation, for example an additional year-to-year adjustment mechanism is required.<sup>14</sup> Nevertheless, when implemented correctly, a WPB can achieve the core characteristics of capitation set out above.

The inclusion of an outcomes-based payment mechanism in WPBs will help mitigate the risk of perverse incentives typically associated with a capitated approach, namely restriction of the number of health and care services provided, particularly services associated with higher costs. Although contracts and regulation will enforce minimum quality, strengthening provider accountability through an outcomes-based payment should enable improvements to health outcomes and service quality.

### 1.1.4. WPB in the context of the National Tariff and GMS/PMS regulations

#### National Tariff Payment System rules

The National Tariff Payment System (NTPS) rules and principles give commissioners and providers broad scope to determine local payment approaches that support the development of more integrated approaches to delivering care.

In developing a WPB to support implementation of an ACO care model, commissioners and providers must be satisfied that their proposed local payment approach complies with the principles and rules detailed in the NTPS local pricing section.<sup>15</sup> In particular, to the extent that a WPB covers services which have national prices under the NTPS, a WPB payment approach may be adopted by the commissioner and provider agreeing “local variations” (ie by agreeing to vary the prices and specifications of the relevant services) in accordance with the NTPS rules.

The provisions of the NTPS cover all NHS-funded healthcare services, except those primary care services where the remuneration of providers is determined by or in accordance with regulations, directions or related instruments under the National Health Service Act 2006 (the 2006 Act),<sup>16</sup> and services which are funded by personal health budget ‘direct payments’. Where payment for primary

<sup>13</sup> WPBs have the same core characteristics as capitated approaches. But while capitated budgets allocate a financial value per person, potentially with different values for patients with different needs, WPBs are based on the historical levels of funding required to care for a population and do not explicitly assign a financial value per head of the population.

<sup>14</sup> Under person-level capitated payment, such adjustment is automatic, based on the movements of people onto/off registered lists.

<sup>15</sup> [https://improvement.nhs.uk/uploads/documents/2017-2019\\_national\\_tariff\\_payment\\_system.pdf](https://improvement.nhs.uk/uploads/documents/2017-2019_national_tariff_payment_system.pdf)

<sup>16</sup> These include core general practice services covered by the GMS or PMS contracts, community pharmacy, dental practice and community optometry.

care services is not determined by the 2006 Act framework, the NTPS rules on local price setting apply. For example, local price setting rules apply to minor surgical procedures performed by GPs and commissioned by clinical commissioning groups (CCGs). Additionally, local authority-funded social care or public health services, including those commissioned under joint commissioning arrangements (a local authority and its NHS partners), are outside the scope of the NTPS.

## GP participation options

Primary medical services and general practice are central to the development of the ACO models. As described in the ACO framework, we envisage three main approaches for voluntary GP participation with different implications for calculating the WPB:

- **Full integration**, where the ACO brings together all primary care services operating under a single WPB. More information about primary care funding streams that can move into the WPB is provided in Section 2.5. The conditions of payment under the ACO Contract will be set out in the contract and supporting documents.
- **Partial integration**, where the ACO Contract excludes primary medical services covered by GMS/PMS contracts. It is supported by contractual arrangements between the ACO and the GPs to achieve operational integration.
- **A 'virtual' model** where separate commissioning contracts are bound together.

### 1.1.5. Essential requirements

Payment design should not be developed in isolation of new care model development as it is an enabling mechanism to support the service transformation. NHS England's guidance in The multispecialty community provider emerging care model and contract framework outlines ten 'essential jobs' to establish a successful MCP.<sup>17</sup> This handbook assumes that local areas have completed these 'essential jobs' or will do so shortly in parallel with WPB development. In addition, local areas will need to consider the following:

#### Data availability and quality

The availability of high quality and up-to-date data on activity, quality and costs is essential for developing, implementing and continual evaluation of a WPB and gain/loss sharing mechanisms, as well as the care models themselves. There is a collective responsibility to ensure that, where appropriate, all parties collect and have access to the information required for such assessments.

#### Population and service scope

The scope of health and care services included in the ACO care model needs to be identified early on. The service scope should be sufficiently wide to prevent cost shifting between care settings and precisely defined so there is no ambiguity about which care activities are in scope. Similarly, the population should be of a sufficient size to support integration of care, minimise risks from random cost variation and incentivise prevention.

#### Shadow test and refine

Where possible, before implementing a WPB, a local area should shadow test the likely financial impacts

<sup>17</sup> <https://www.england.nhs.uk/wp-content/uploads/2016/07/mcp-care-model-frmwrk.pdf>

<sup>18</sup> Whole Population Budgets: Shadow testing webinar is available at: <https://improvement.nhs.uk/events/capitated-whole-population-budgets-shadow-testing-webinar/>

and necessary payment operations. This involves scenario testing to understand and refine any incentives and to evaluate the appropriateness of the baseline, forecast and gain/loss sharing mechanism.<sup>18</sup>

## **Provider-to-provider payments**

A clear set of rules should be locally agreed to determine payment to any subcontractors and/or distribution of the WPB between the different parties in the ACO care model, including any gains/losses under a sharing arrangement. Any provider-to-provider payments should continue to support the objectives of the care model.

### **1.1.6. Commissioner considerations**

#### **Agreeing the WPB**

Commissioners must have regard for their statutory duties and other obligations.<sup>19,20,21</sup> As such, they need to assure themselves that the WPB for an ACO is consistent with appropriate distribution of funding between the ACO and all other services they are accountable for commissioning.

In particular, commissioners, in conjunction with key stakeholders, need to consider system sustainability, ensuring that the multiyear WPB is both affordable and covers efficient costs. They need to review the planned WPB in the context of pressures across the wider local health and care system, forecasting demand and spend on services within and outside the scope of an ACO in parallel, and ensuring consistency with local sustainability and transformation plans (STPs).

This review needs to consider the national model for oversight and regulation of new care models (in development) and any requirement to mitigate risks, including those set out by the Integrated Support and Assurance Process (ISAP).<sup>22</sup> NHS Improvement has recently published the Single Oversight Framework<sup>23</sup> to support NHS providers in attaining and maintaining the standards required to meet their regulatory obligations, including during the transition to new care models.

Where the estimated WPB (and therefore the ACO care model) is considered unaffordable or unsustainable in the context of the overall local STP and commissioning plans, commissioners need to consider what adjustments they must make to maintain system sustainability. These are likely to include changes to the ACO care model or its scope, as well as potential changes to other local commissioning plans. In all cases commissioners will need assurances that delivery of the contract obligations within the WPB value is achievable for the population and services in scope of the ACO, and as such local areas need a shared understanding of how any savings will be realised.

#### **Patient choice**

Patient choice is a key feature of NHS care. The NHS Choice Framework<sup>24</sup> sets out the choices patients can make concerning their care. When patients covered by the WPB choose to receive services from providers outside the ACO, these providers will need to be reimbursed from the WPB and appropriate arrangements will need to be set up for this. ACO funding will also need to be adjusted for patients choosing to register with (or leave) the ACO as their primary care provider during the contract period.

19 <https://www.england.nhs.uk/about/gov/equality-hub/legal-duties/>

20 NHS Act 2006: <http://www.legislation.gov.uk/ukpga/2006/41/contents>

21 Health and Social Care Act 2012: <http://www.legislation.gov.uk/ukpga/2012/7/contents/enacted>

22 <https://www.england.nhs.uk/resources/resources-for-ccgs/#isap>

23 [https://improvement.nhs.uk/uploads/documents/Single\\_Oversight\\_Framework\\_130916.pdf](https://improvement.nhs.uk/uploads/documents/Single_Oversight_Framework_130916.pdf)

24 <https://www.gov.uk/government/publications/the-nhs-choice-framework/the-nhs-choice-framework-what-choices-are-available-to-me-in-the-nhs>

## Integrated personal commissioning

The integrated personal commissioning (IPC) programme supports integration and personalisation of services by joining up health, social care and other services at the level of the individual.<sup>25</sup> Part of this programme includes expanding the use of personal health budgets (PHBs) to give a small number of people with the highest care needs, such as those with long-term health conditions or disabilities, more choice and control over how money is spent on meeting these needs.

As PHBs are calculated at an individual level, how these funds will be administered in relation to a WPB needs to be considered and what is agreed needs to be recorded as part of the ACO Contract. Funding will flow accordingly.

## Procurement and financial assurance processes

The Procurement and assurance approach<sup>26</sup> provides commissioners with information on the likely considerations for the procurement of an ACO Contract. It sets out the regulatory landscape, provides an update on the development of the ISAP by NHS England and NHS Improvement, and sets out a number of common principles and considerations to support local processes.

## 1.2. Overview of Whole Population Budget development

NHS England and NHS Improvement have been working with a number of MCP and PACS vanguards to develop the method for calculating and implementing a WPB. Each vanguard currently has a different scope of health and care services within its new care model. This variation has provided valuable lessons for inclusion in this handbook.

The main stages in developing a WPB are:

- calculating the WPB baseline
- estimating WPB values for future years
- converting estimated WPB values to contract values for each year in a contract
- introducing gain/loss sharing arrangements.

An overview of each of these stages is given below, with subsequent chapters describing them in detail.

### 1.2.1. Calculating the WPB baseline

The WPB baseline is the money available to an ACO to fund services for the target population in the first year of the contract. It is likely to be calculated, at least initially, using data on commissioner spend because of limitations in the understanding and transparency of provider costs, particularly at patient level and in community settings.

The baseline spend is currently spread across a number of separate contracts for different care settings. Chapter 2 sets out the method for disentangling and then amalgamating the necessary data to calculate the WPB baseline. The starting point is a clear definition of both the scope of services and the population covered by the ACO.

25 <https://www.england.nhs.uk/healthbudgets/wp-content/uploads/sites/26/2016/05/ipc-emerging-framework.pdf>

26 <https://www.england.nhs.uk/publication/procurement-and-assurance-approach-document-4>

### 1.2.2. Estimating WPB values for future years

To facilitate system transformation and realise health and cost benefits from this, it is likely that initial ACO contracts will have a duration of up to 10 years.<sup>27</sup> Organisations need to forecast the WPB for each year of the contract and to do this using transparent mechanisms and assumptions about future economic factors. These forecasts, to be included in the contract, will facilitate investment by setting provisional indications at the procurement stage of what the WPB might be in each year of the contract.

Chapter 3 provides guidance to help commissioners (and providers where appropriate) to forecast the WPB value for the duration of the contract, which may include planned scope changes as specified in contracts. These forecast values will take account of expected:

- future cost and activity pressures, for example anticipated changes in population size, demographics as well as inflation in health and care provision costs (as reflected in national planning assumptions)
- provider efficiencies, for example annual efficiency targets on unit costs in acute settings
- funding requirements associated with implementing the new care model plus any efficiencies over and above those expected of the NHS more broadly as the new care model scales up.

Commissioner allocations act as an overall constraint on the level of funds available for health and care services in the local area. This together with wider commissioning and service plans for the local health economy and local STPs must be recognised in estimating the WPB value for future years.

### 1.2.3. Converting estimated WPB values into contract values

The baseline and forecast values are the basis for agreeing the WPB contract values. But this needs to recognise that it will not be possible to, for example, accurately forecast population movements for the full contract period. Commissioners and providers will need to agree a mechanism to adjust the WPB contract value periodically so that it continues to reflect the needs of the population in scope of the ACO and to allow the use of up-to-date or improved data. Chapter 4 provides technical guidance on this.

Commissioners should refer to the guidance in the *Procurement and assurance approach*<sup>28</sup> to identify where changes to service or population scope would trigger a renegotiation of the ACO contract itself, not just an adjustment to the WPB value in the contract.

### 1.2.4. Improvement Payment Scheme

The implementation of an outcomes-based payment scheme is intended to signal and encourage change across a range of priority areas. A national IPS will help mitigate the well-documented risk that introduction of a WPB may result in rationing of access to and/or reduced quality of care. The IPS payment constitutes a portion of the contracted integrated budget value and is paid upon delivery against targets for agreed metrics. In addition to the paid for element of the quality-incentive scheme, we will set out a range of outcome metrics against which ACOs' performance will be published. Chapter 5 provides more information about the IPS.

27 <https://www.england.nhs.uk/wp-content/uploads/2016/07/mcp-care-model-frmwrk.pdf>

28 <https://www.england.nhs.uk/publication/procurement-and-assurance-approach-document-4>

### **1.2.5. Gain/loss sharing**

The introduction of a gain/loss sharing mechanism in addition to a WPB will further support the transformation of health and care services. Chapter 6 describes how gain/loss sharing can help to:

- better align financial incentives across services not covered by the WPB
- manage the transition and impact of utilisation risk to providers over time.

## 2. Calculating the Whole Population Budget baseline

### 2.1. Overview of Whole Population Budget development

The Whole Population Budget (WPB) baseline is the money available to an ACO to fund services for the target population in the first year of the contract. It is the starting point for forecasting the WPB over the contract term (see Chapter 3).

This chapter sets out a method for calculating service-level baselines across different care settings and contract types. It is structured around two steps commissioners and providers should undertake collaboratively to set a WPB baseline:

- **Calculate the current commissioner spend at service level** for all in-scope services. This will depend on the type of existing contract and payment approach for services in scope of the ACO:
  - activity-based payment (Section 2.3)
  - block contracts (Section 2.4)
  - primary care contracts (Section 2.5)
  - social care packages (Section 2.6).
- **Convert the current spend to the WPB baseline** by adjusting it to reflect in-year efficiency and inflation assumptions, efficient provider costs and planned service investment, or to correct data anomalies. This step (Section 2.7) focuses on adjusting the most recent year's spend to derive the WPB baseline, as distinct from estimating WPB values for future years of the contract (Chapter 3).

#### 2.1.1. Foundational activities

While it is potentially more straightforward to calculate a WPB baseline than to forecast WPB values for the duration of the contract or to structure gain/loss arrangements, stakeholder engagement is essential from the start through to eventual procurement. Any conflicts of interest for providers in sharing cost data with commissioners during this process need to be handled appropriately. The baseline calculation process can be used to build trust between organisations and to set up communication structures that will be beneficial when other aspects of the WPB are tackled.

At the outset, commissioners in areas seeking to develop a WPB should decide which organisations to involve in calculating the baseline – at a minimum, a commissioner and provider(s) committed to engaging collaboratively and transparently. Governance arrangements should be defined to support joint working and effective decision-making, and to provide a forum for information sharing.

### 2.2. Addressing likely data challenges

The ideal WPB baseline should reflect the efficient costs of delivering services. However, a lack of appropriate data means that determining efficient costs is challenging in most localities. The sum of commissioner spend on in-scope services is therefore the recommended pragmatic starting point for calculating the baseline value. The resulting baseline should be adjusted to better reflect efficient costs through commissioner and provider engagement (Section 2.7).



If data is not available for the current year (ie the year before the contract is due to start), there are two broad options for estimating the baseline for a service:

- Adjust the **previous year's spend** data to reflect the current year. As a minimum, adjustments need to be made for three factors:
  - expected spend change, including for:
    - tariff alterations between the previous and current year for services paid on activity
    - inflation
    - efficiency savings, incorporating quality, innovation, productivity and prevention (QIPP) targets
  - expected service investment in the current year, including to establish the new care model
  - expected change in population size or healthcare needs.
- Use a **partial sample of the current year's spend data** and extrapolate – accounting for seasonality – to the full year. The accuracy of this approach is limited if only a few months of the current year's data is available. It also relies on the assumption that historical seasonality is a good proxy for the current year's seasonality.

Whichever option is used, checking the baseline against several years of historical data will give assurance that it is not an outlier.

## 2.3. Service-level spend calculated from activity-based payments

Payment by activity covers the majority of acute healthcare in hospitals, with national prices for most admitted patient care, outpatient attendances, A&E and some outpatient procedures. National prices are also used in some other areas of the health system.

For **acute services**, the Secondary Uses Service (SUS) database can be used to isolate activity related to the intended population (using the registered GP field) or services (using the specialty, diagnostic or procedure fields, depending on how best to identify the service). In turn, the annual value of the filtered activity can be calculated on current prices.

For **other services** reimbursed by activity (eg some mental health secondary care), an analogous method is preferred:

- isolate activity data for the intended population at the same level of detail as the tariff
- attach a value to the service using current prices.

In practice, trusted data at the right level of detail may not be available. A programme of action to collect the right data is the best course to improve accuracy of the service value calculation. In the near term, there are some practical workarounds (Table 1).

**Table 1: Workarounds for data challenges**

Data challenge	Possible workaround
<b>Activity data does not have the right fields to isolate in-scope services</b> Example: the total spend on a care setting is known, but it provides some services which are out of scope of the care model	Use the methods in Section 2.4 (service-level spend on in-scope services within block contracts) to apportion spend
<b>Activity data does not have the right fields to isolate the in-scope population</b> Example: the value of a service is known for a population which includes patients registered with GP practices that are not part of the ACO	Estimate the value of the services provided to the in-scope population by: <ul style="list-style-type: none"><li>• <b>weighted carve out:</b><ul style="list-style-type: none"><li>- identify disease drivers for use of a service</li><li>- use disease prevalence data at the GP practice level to apportion use of a service</li></ul></li><li>• <b>population proportional carve out:</b> assume uniform use of activities across the patient population to calculate the apportionment (lowest accuracy).</li></ul>

## 2.4. Service-level spend on services within scope of block contracts

Identifying the service-level baseline from block contracts is straightforward if the services and population covered by the block contract matches the scope of the WPB. If not, it is necessary to disaggregate the block contract spend to isolate the share attributable to the in-scope services and population.

In this section we describe how to allocate commissioner spend to the patient level, to estimate spend for the relevant GP lists and for the services within scope of the WPB. Some of the methods use provider costs, but the resulting figures are estimates of commissioner spend per unit of activity. The terms 'commissioner spend' and 'block contract amount' correspond to provider revenue, as distinct from 'provider costs', which are the costs incurred by the provider in the delivery of care.

### 2.4.1. Overview of the approach to block contracts

The proposed approach has four steps (see Figure 1):

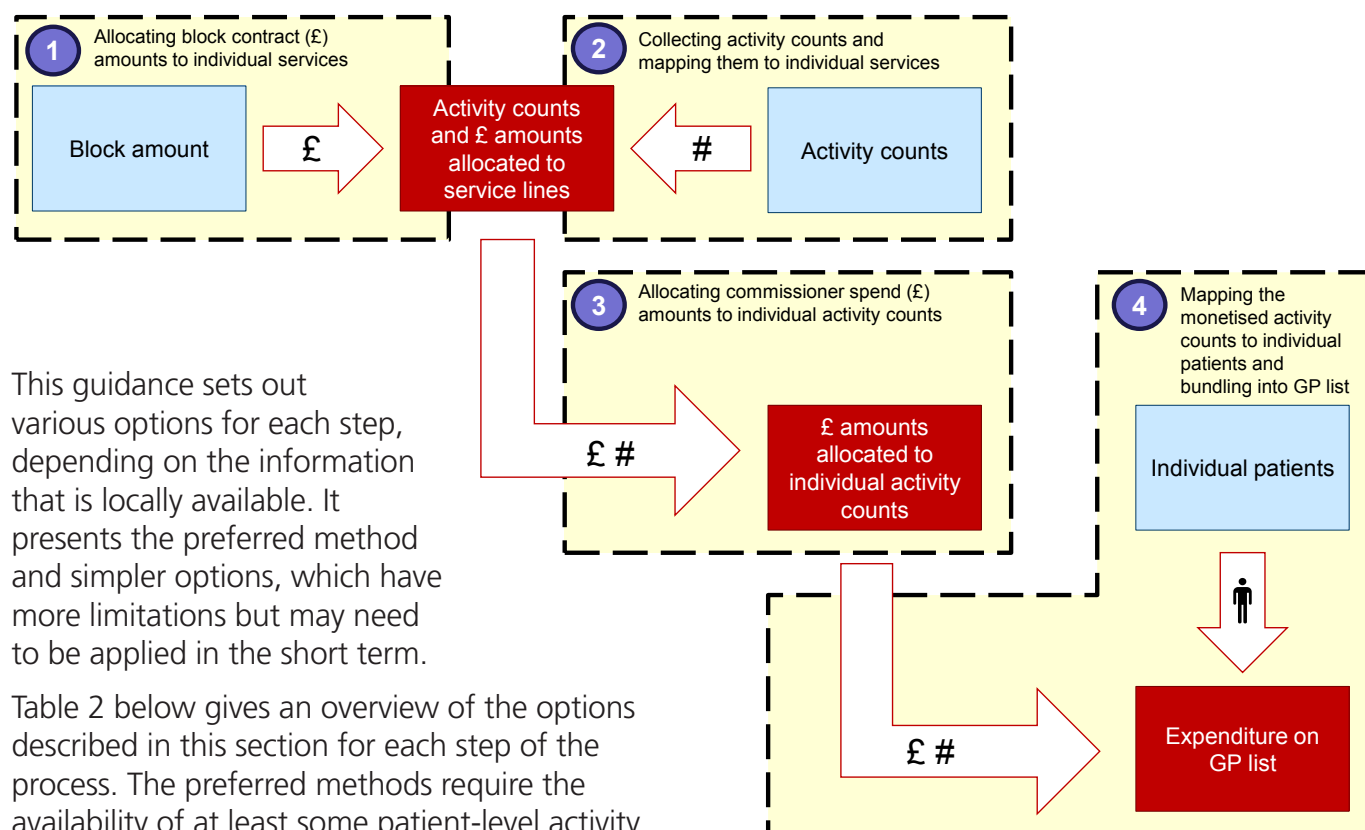
**Step 1:** Break up the block contract (£) amounts to allocate them to individual service lines.

**Step 2:** Collect activity counts and map them to the service lines, so that both activity counts and commissioner spend (£) amounts are allocated to the various service lines.

**Step 3:** For each service line, allocate the estimated commissioner spend (£) amounts to individual activity counts, to produce monetised activity counts.

**Step 4:** Map the various monetised activity counts to individual patients (or patient groups) and bundle them into GP lists, to produce estimates of commissioner spend on specific GP lists.

**Figure 1: Estimating commissioner spend per patient (or GP list) – overview of proposed approach**



This guidance sets out various options for each step, depending on the information that is locally available. It presents the preferred method and simpler options, which have more limitations but may need to be applied in the short term.

Table 2 below gives an overview of the options described in this section for each step of the process. The preferred methods require the availability of at least some patient-level activity counts (for the relevant commissioner or provider).

Three general principles should govern the process:

- **Constructive engagement between providers and commissioners:** information sharing between commissioners and providers will enable more accurate estimates (and can prevent duplication of work). If WPB assumptions can't be validated with providers before procurement, it is essential that the WPB calculation is appropriately transparent to all potential bidders and that the WPB value is assured by provider due diligence.
- **Reconciliation of commissioner spend and provider cost:** the commissioner spend per patient (or GP list) estimates resulting from this exercise should be compared to provider cost figures where available. Possible discrepancies should be considered when using these estimates, particularly where individual service line values in existing block contracts do not align with the cost of the service lines and providers are thought to be making internal cross-subsidies across the scope of the current contract.
- **Regular review of accuracy:** local areas should regularly review the commissioner spend, provider cost and activity information used throughout the process, to ensure that it is accurate and up to date. Where it is not, local areas should seek to update the information or, if this is not possible, take into account the potential inaccuracy this may introduce into the commissioner spend per patient (or GP list) estimates (especially as these inaccuracies may be magnified if they are used to project a payment amount over multiple years).<sup>29</sup>

<sup>29</sup> When using these estimates, local areas could for instance carry out a scenario modelling exercise to assess the impact of potential inaccuracies, and introduce gain/loss sharing arrangements and revision clauses into the relevant contracts.

We do not provide detailed guidance on how to:

- design and define units of activity/currencies, although the high-level principles to follow are given
- link person-level information between providers (once the allocation has been done).

**Table 2: Overview of methods to disaggregate block contracts**

Step	Preferred option	Second option	Last option
<b>1</b> Allocating block contract (£) amounts to individual services Section 2.4.2	Block contract schedule with accurate service line breakdown of £ values	Using provider service line costs to allocate commissioner spend	No allocation of the block contract (£) amount to service lines
<b>2</b> Collecting activity counts and mapping them to individual services Section 2.4.3	Information following nationally determined standards	Other locally collected activity information	
<b>3</b> Allocating commissioner spend (£) amounts to individual activity counts Section 2.4.4	Using available estimates of relative resource consumption of different types of activity as weightings to allocate commissioner spend	Using nationally collected or estimated information (eg reference costs) as weightings to allocate commissioner spend	No differentiation between types of activity (calculating average commissioner spend across total activity count)
<b>4</b> Mapping the monetised activity counts to individual patients and bundling them into GP lists Section 2.4.5	Using the patient identifiers	Mapping exercise based on available information on care needs and diagnostics, and input from care professionals	

This guidance does not cover the two circumstances where commissioner spend per GP list (or per patient) is readily available and therefore the service-level baseline is straightforward to calculate:

- block contracts cover services for a specific GP list (unless local areas are looking to estimate commissioner spend per patient)
- unit cost figures are available for services.

#### **2.4.2. Step 1: Allocating block contract (£) amounts to individual services (see Figure 2)**

The level of detail in the available data will vary across local care economies. In this section we describe possible approaches where:

- contracts contain an accurate service line breakdown of the total amount (preferred option)
- there is no accurate service line breakdown in the contract, but provider service line cost information is available (second option)
- there is no information that allows a breakdown of the total contract (£) amount by service line (last option).

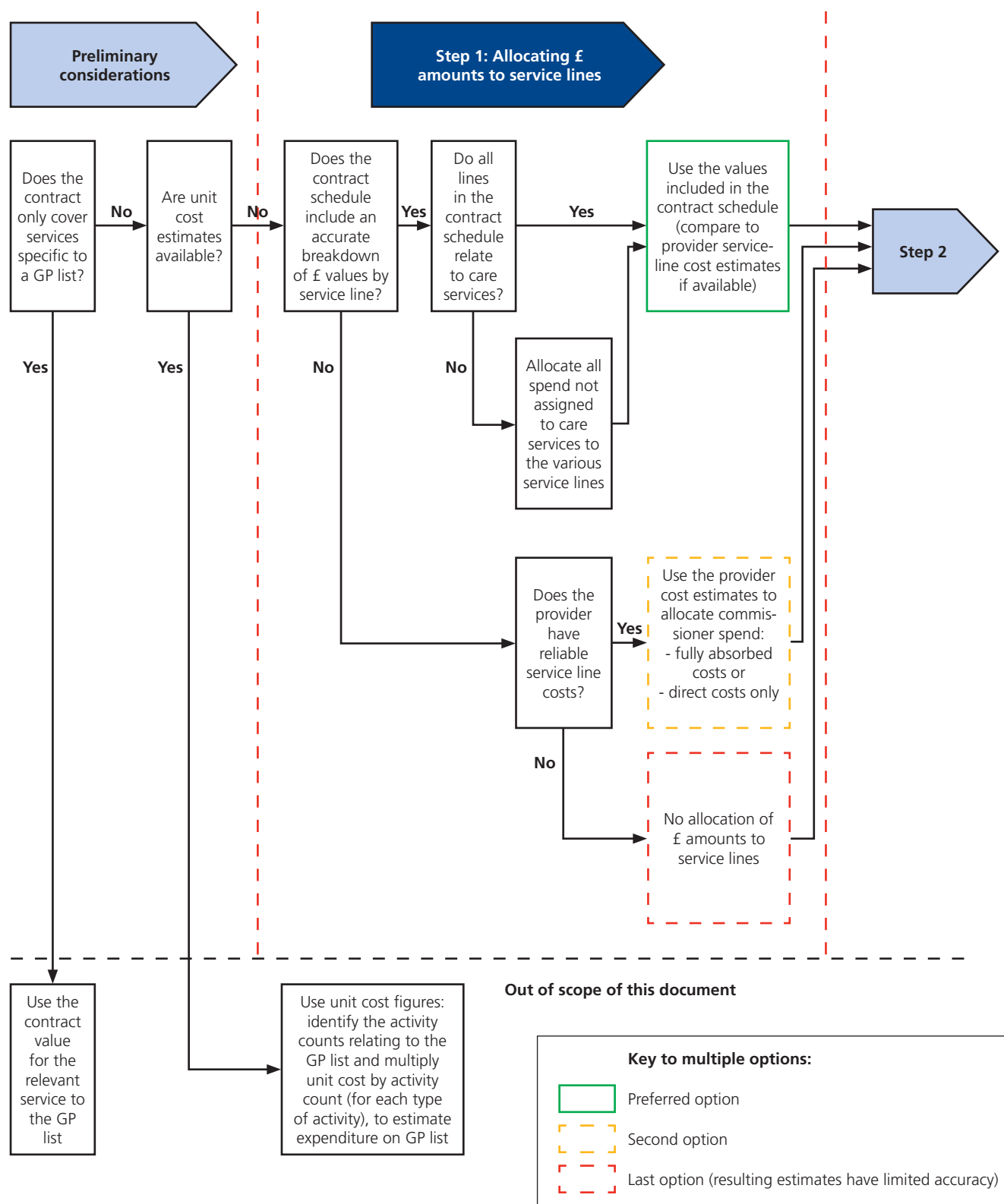
Preferred option for Step 1: Block contract schedule with accurate service line breakdown of £ values

If the contract schedule includes a breakdown of £ values by service line, local areas should first consider whether these values reflect the actual spend by service line. If this is the case, they should use the information in the contract schedule to allocate the total contract (£) value to the various service lines (if not, they should seek to use the second option described below).

In addition, if the contract schedule includes lines that do not relate directly to a type of care service (eg 'Admin' in the example in Figure 3 below), the £ values assigned to these lines should be reallocated to the various care services.

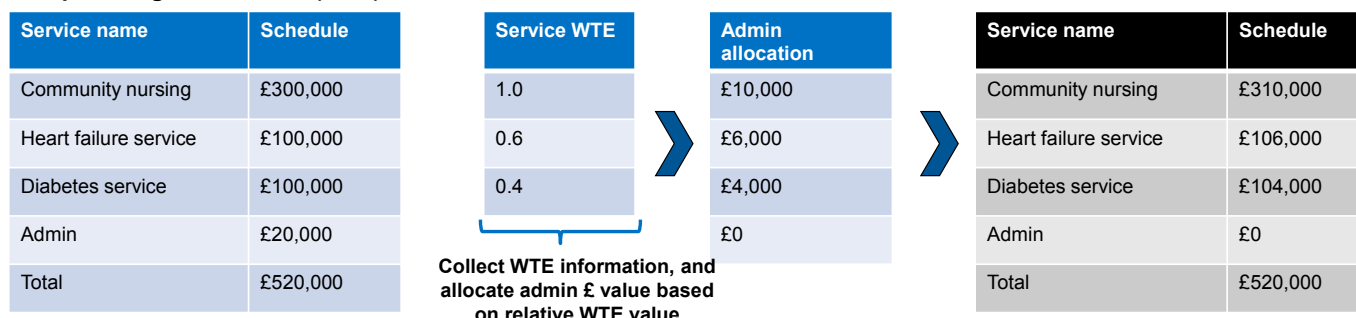
Figure 3 below provides an example of three possible methods to allocate admin costs across the care services contained in a block contract. The most accurate method uses a cost driver (1), following the steps listed in Box 1 below. Alternatively, local areas could use the contract schedule to allocate commissioner spend based on the relative £ amounts assigned to each service (2), or allocate admin spend homogeneously across all services (3).

**Figure 2: Step 1 – overview**

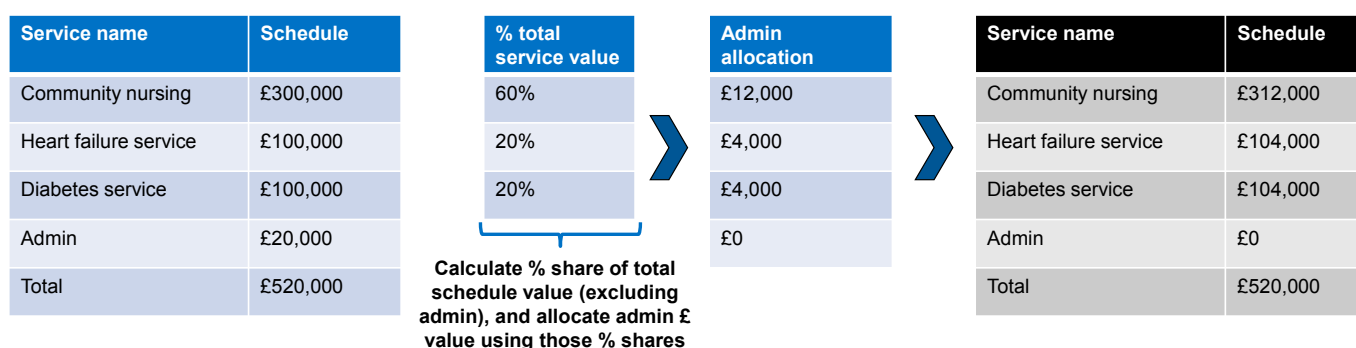


**Figure 3: Allocating £ amounts not assigned to a service line to the various care services – hypothetical example (allocating ‘Admin’ value)**

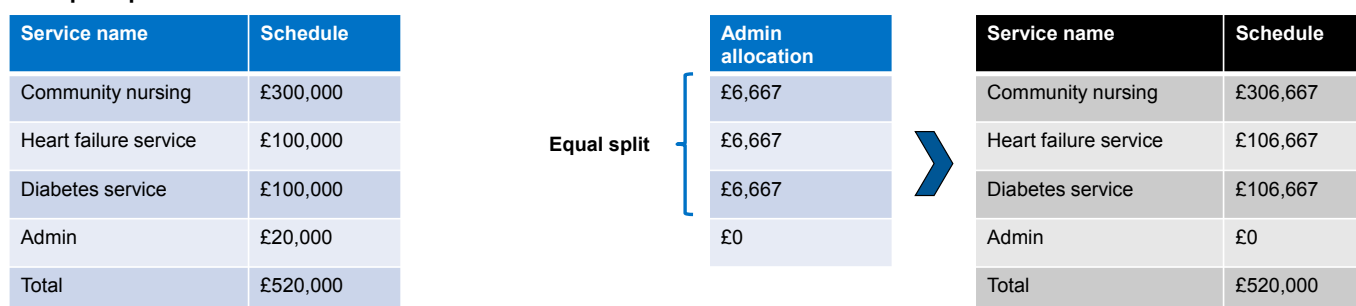
### 1. Split using a cost driver (WTE)



### 2. Schedule value split



### 3. Equal split



If in addition to an accurate block contract schedule broken down by service line, local areas have access to provider service line cost information, they should sense check the figures in the block contract against the relative values (rather than absolute £ values) of the provider costs across the different service lines.

If data are not available to make an accurate service line breakdown, alternative approaches are presented below.



### Box 1: Using a cost driver

To reallocate £ values that are not directly associated with a care service in the contract schedule, local areas should ideally identify an appropriate cost driver<sup>30</sup> by following three steps:

1. Identify the cost object (ie the unit to which local areas wish to allocate a £ value, for instance a care service).
2. Make a list of potential cost drivers (ie factors that could impact the cost of providing the service, for instance the duration of each consultation).
3. Determine the relationship between the potential costs drivers and the cost of the cost object (for instance, cost is proportional to the duration of the consultation, and each hour costs £20).

#### *Second option for Step 1: using provider service line costs to allocate commissioner spend*

If the block contract schedule does not include a breakdown by service line, or includes one that is thought to be inaccurate, local areas should use provider service line cost estimates to allocate commissioner spend to the various service lines. There are two possible alternative methods to the preferred option, the first of which is more accurate than the second but requires more information. These methods use:

- provider's fully absorbed costs
- provider's direct service costs only.

Where the block contract schedule does give a breakdown of £ values by service line but this is deemed inaccurate (or out of date), local areas could use these methods to refresh the values in the schedule.

Note that, in addition to constructive engagement between commissioner(s) and provider(s), these methods rely on commissioners having access to provider cost information.

**Using provider's fully absorbed service costs** If possible, local areas should engage with providers to get information on their fully absorbed costs (ie total costs absorbed at a service line level). If non-patient care costs (such as education and training, commercial activities, and research and development) can be separated out, they should be excluded from the calculation. Otherwise, all provider costs should be allocated to all patient services.

Commissioners should seek assurance from providers that the costs were calculated by following NHS Improvement's Approved costing guidance and, in future, the mandatory costing standards.<sup>31</sup>

**Using provider's direct service costs only** If there are concerns about the accuracy of the indirect or overhead costs, it may be more appropriate to use provider's direct costs only to allocate commissioner spend to service lines.<sup>32</sup> This may be because direct costs are easier to identify at a service level. Figure 4 below shows an example of this method.

30 "A cost driver can be described as any factor that causes a change in the cost of an activity": [http://www.cimaglobal.com/Documents/ImportedDocuments/ABM\\_techrpt\\_0401.pdf](http://www.cimaglobal.com/Documents/ImportedDocuments/ABM_techrpt_0401.pdf), p4

31 [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/404708/Approved\\_costing\\_guidance\\_-\\_17\\_Feb\\_2015.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/404708/Approved_costing_guidance_-_17_Feb_2015.pdf)

32 For definitions of direct, indirect and overhead costs please see HFMA's clinical costing standards: <http://www.hfma.org.uk/costing/standards>

**Figure 4: Allocating commissioner spend (£) amounts to service lines using provider direct costs – hypothetical example**

Collect provider direct costs, and calculate % of total for each service line

**1** (Direct costs for service / total direct costs)

Services	Contract value	Provider direct costs (£)	Provider direct costs (% total)	Estimated commissioner expenditure
Diabetes care		£200,000	31%	£153,846
Dementia care		£350,000	54%	£269,231
Speech & language therapy service		£100,000	15%	£76,923
<b>Total</b>	<b>£500,000</b>	<b>£650,000</b>	<b>100%</b>	<b>£500,000</b>

**2** Calculate estimated commissioner expenditure  
(direct costs (% total) x total contract value)

Local areas should be aware that if they use direct service costs only as weightings, the resulting commissioner spend estimates for services with disproportionately high/low levels of indirect and overhead costs may be inaccurate. For instance, a service with a higher than average proportion of direct costs (as a share of total costs) will be assigned an overestimated commissioner spend value.

*Last option for Step 1: No allocation of the block contract (£) amount to service lines*

If neither of the other two methods above can be used, local areas should move on to Step 2 without allocating the block contract (commissioner spend) amount to individual service lines.

Not being able to allocate the block contract (commissioner spend) amounts to the relevant service lines will make Step 3 of this guidance (in particular the preferred option) more difficult.

### 2.4.3. Step 2: Collecting activity counts and mapping them to individual services (see Figure 5)

Once Step 1 has been completed and spend (£) amounts have been assigned to service lines (unless the last option for Step 1 is the only one available), local areas should gather the available activity data. They should then map the activity counts to the individual service lines, so that both activity counts and commissioner spend (£) amounts are allocated to the different types of services.

Local areas can use a number of sources of activity information, from information following national specifications to data reported locally to the commissioners or held by the relevant providers.

*Preferred option for Step 2: Collecting activity information following nationally determined standards*

Local areas could use the activity information sources listed below, which all follow national standards for acute, community and mental health services, if they cover the relevant services:

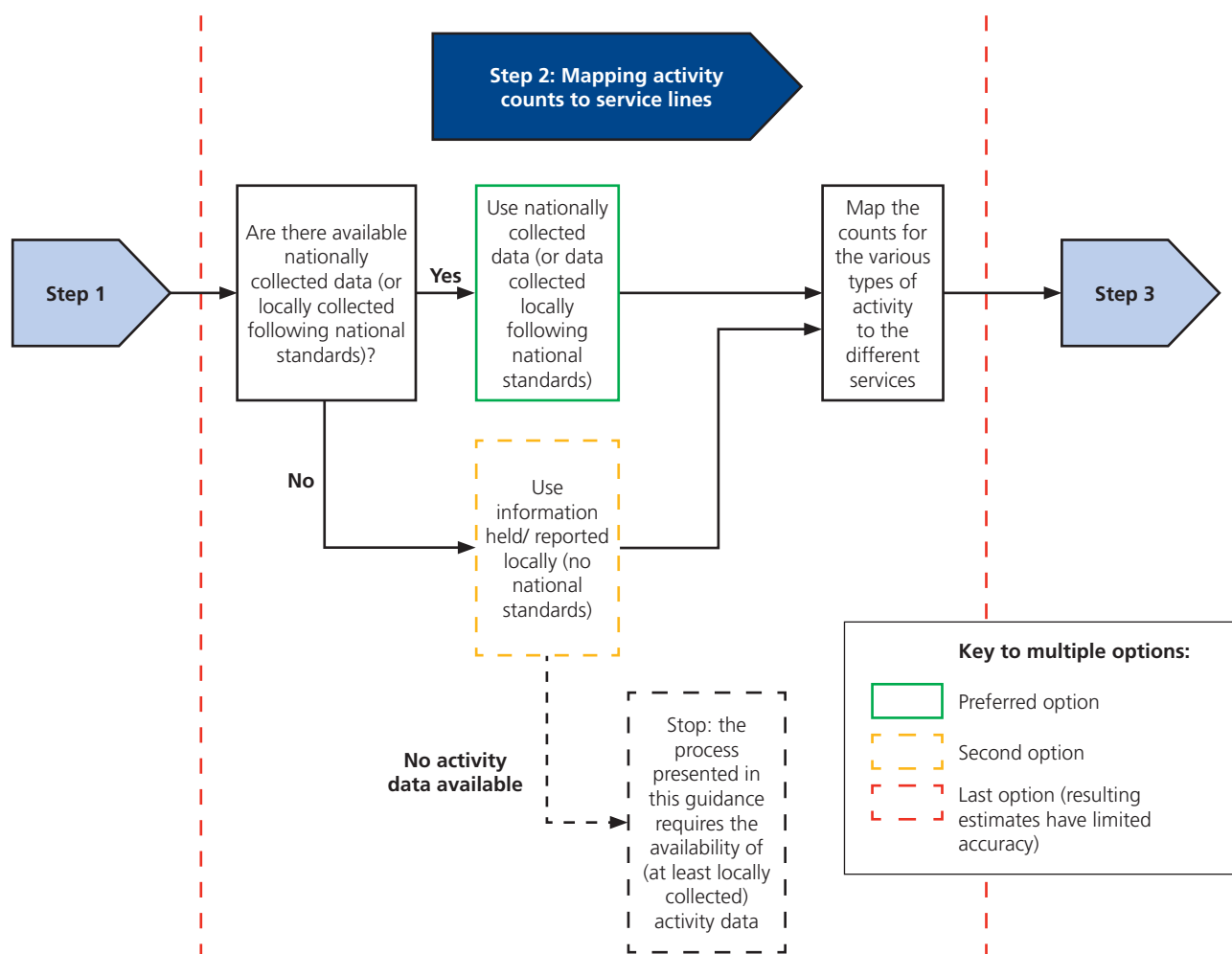
- services provided in acute settings are reported at the episode level to the commissioner, and submitted to the central SUS
- reference costs<sup>33</sup>
- Community Information Data Set (CIDS)<sup>34</sup>
- mental health clusters
- Mental Health and Learning Disabilities Data Set<sup>35</sup>: the information submitted by providers as part of the dataset should include activity data (inpatient, outpatient, bed days, etc) at a patient level. However, this information is only available at an aggregated level when sent back to local commissioners, so engagement between commissioner and provider (alongside appropriate information governance arrangements) will be necessary to share and use the patient-level information.

33 <https://www.gov.uk/government/collections/nhs-reference-costs>

34 <http://www.hscic.gov.uk/comminfodataset>

35 <http://www.hscic.gov.uk/mhldds>

**Figure 5: Step 2 – overview**



### *Second option for Step 2: Other locally collected activity information*

The preferred option cannot be used for services not covered by the activity data sources listed above. In some cases, service line reports to the commissioner (or information collected for local dashboards) will include activity data that could be used for this exercise. For instance, in South Manchester these reports provide useful information and detail for each service, counts for each type of attendance (first, follow-ups, etc), for each 'location' (home, clinic, phone), as well as the number of referrals for each 'priority type' (routine, urgent, two-week rule, A&E).

In addition to the data regularly reported to their local commissioner(s), providers may collect activity information at the patient level for their own performance and risk management purposes.

Local areas should assess the appropriateness of collected activity data (the key principles to follow are given in Box 2). This will help them to identify potential sources of inaccuracies in their estimation of commissioner spend per patient (or GP list).

## Box 2: Assessing the appropriateness of the available activity data

Although commissioners and providers will likely be constrained by the data available, they can assess the appropriateness of this data by checking it has the following features:

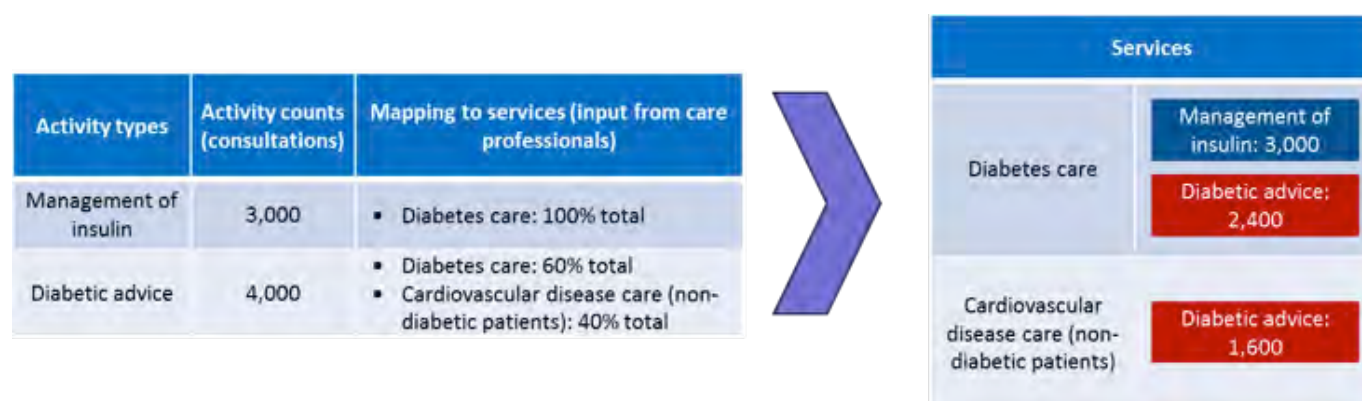
- the activity data used should be relevant from a clinical point of view and/or the point of view of the patient or service user
- the different units of activity should be mutually exclusive (ie no overlap), and collectively exhaustive (ie no activity is unaccounted for)
- each type of activity should be homogeneous (especially in terms of inputs): activity X for patient A should be very similar to that for patient B
- units of activity should be consistent across the local providers. Ideally, there should also be consistency across local and non-local providers, which would make benchmarking possible
- the accuracy of information should be easy to assess, which would be the case, for example, if it originates from audited accounts or the commissioner has visibility of how the data was collected
- ideally, time series should be available, which would enable setting of baselines and the identification of outliers (eg years of very high/low volumes of certain types of activity)
- data should be at the patient level, ideally with a patient identifier as this will enable the direct allocation of monetary values to specific (types of) patients. At a minimum, the activity data used in this exercise should be at the patient-service level
- it may also be easier if the units of activity used by the commissioners are consistent with those used by the providers for their own costing purposes.

### *Mapping activity counts to service lines*

Once activity data has been collected, commissioners and providers should work together to map it against the various service lines used in Step 1 (described in Section 2.4.2 above).

Where types of activity are specific to individual service lines, they should be mapped accordingly (eg management of insulin consultations assigned to diabetic care). Where this is not the case, local areas should seek input from clinicians and staff members with the relevant knowledge to apportion the activity counts between the relevant services (eg 60% of dietetic advice is provided for diabetic care and 40% to non-diabetic patients with cardiovascular disease). Figure 6 below presents an example.

**Figure 6: Mapping out activity counts to service lines – hypothetical example**



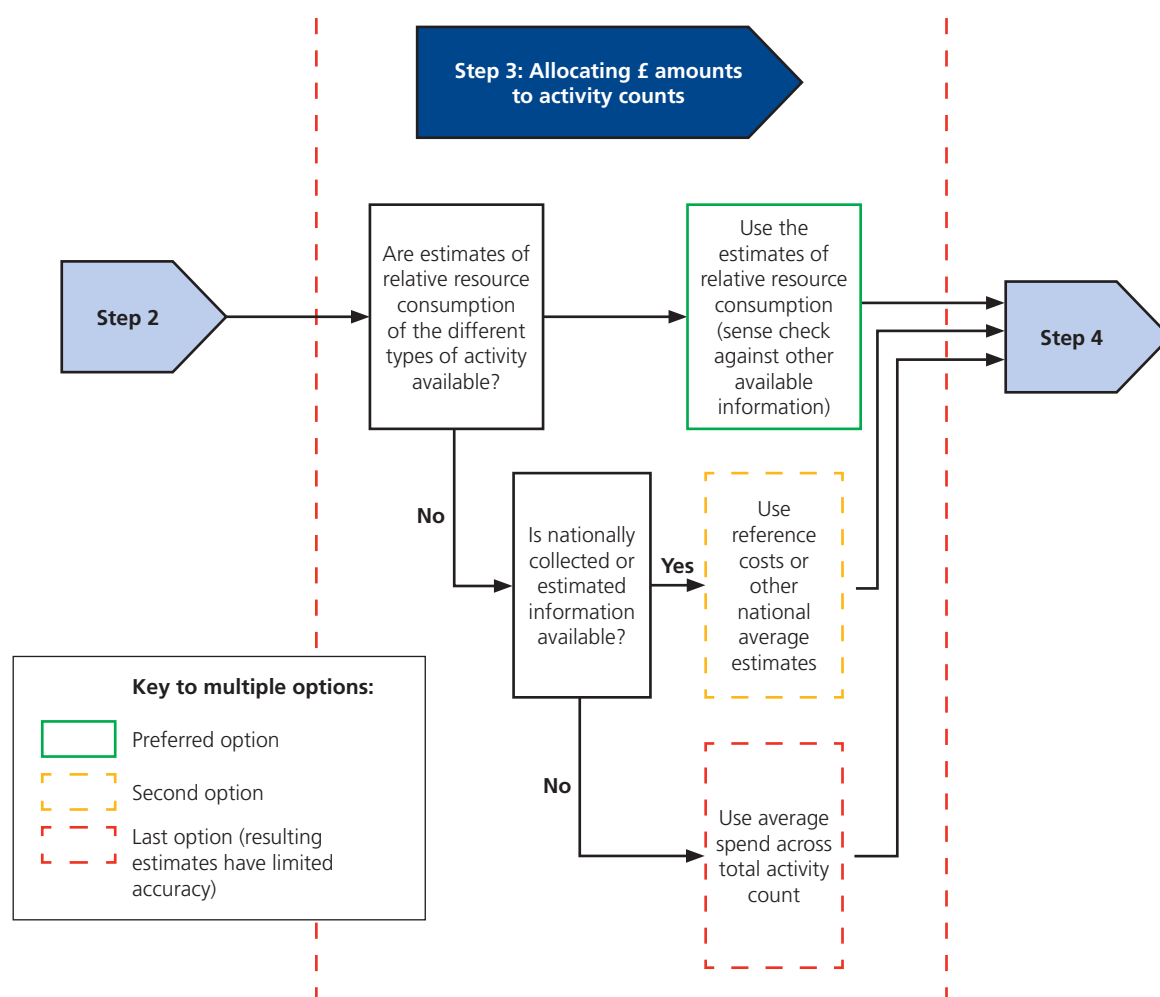
#### 2.4.4. Step 3: Allocating commissioner spend (£) amounts to individual activity counts

Once block contract (£) amounts and activity counts have been assigned to the individual service lines, local areas can use this information to allocate commissioner spend to individual activity counts. There are three options (see Figure 7 below):

- using available estimates of relative resource consumption of different types of activity as weightings to allocate commissioner spend. This can be based on a time and motion study (sample observation of resource consumption) (preferred option)
- using nationally collected or estimated information (eg reference costs) as weightings to allocate commissioner spend (second option)
- no differentiation between types of activity (calculating average commissioner spend across total activity count) within the service line (last option).

The preferred and second options use weightings to estimate a commissioner spend value specific to each type of activity, whereas the last option does not differentiate between activity types and is therefore less accurate.

**Figure 7: Step 3 – Overview**





The options above range from an analytically robust yet resource intensive approach (time and motion studies) through to an approximate method which requires only minimal resource (no differentiation between activity types). Sites should decide which option to pursue based on the degree of rigour required at this step to produce an accurate baseline, balanced against local resource availability.

*Preferred option for Step 3: Using available estimates of relative resource consumption of different types of activity as weightings to allocate commissioner spend*

In the absence of actual provider cost information (a case which would otherwise be outside the scope of this handbook; see Figure 2), the preferred option is to use estimates of the relative resource consumptions of the different types of activity (eg home visits versus clinic consultations) to allocate the service line commissioner spend estimates to individual activity counts. To do this, local areas could observe the resource utilisation of different types of activity for a sample of patients/service users in a time and motion study.

Time and motion studies look to allocate commissioner spend to each activity count by assigning cost weightings to the different types of activity. This requires input from the relevant provider.

Figure 8 below presents a hypothetical example of the use of cost weightings to allocate commissioner spend to different types of activity counts. This approach applies to any of the methods used to produce cost weightings described below.

**Figure 8: Using cost weightings to allocate commissioner spend to different types of activity counts – hypothetical example**

Calculate 'weighted activity' per service line

**1** (Activity x cost weighting)

Drug services (non-admitted)	Activity provided	Contract value	Cost weighting	Weighted activity	Cost activity weighting	Estimated commissioner expenditure
Adult home visit	14,945		1.0	14,945	60%	£1,806,501
Adult clinic appointment	3,789		0.8	3,031	12%	£366,401
Child & adolescent home visit	211		2.5	527.5	2%	£63,762
Child & adolescent clinic visit	2,105		3.0	6,315	25%	£763,336
<b>Total</b>	<b>21,050</b>	<b>£3,000,000</b>		<b>24,819</b>	<b>100%</b>	<b>£3,000,000</b>

Calculate cost-activity weighting  
(Weighted activity / total weighted activity)

**2**

**3**  
Calculate estimated commissioner expenditure per service  
(Total Contract Service Value x Cost-activity weighting)



## Time and motion study (sample observation of resource consumption)

Local areas could use the observed resources used by a sample of patients/service users for each type of activity (eg time with relevant staff) as weightings. These weightings will then allow local areas to allocate different commissioner spend (£) values to different types of activity, following the same process as in Figure 8 above.

The resulting estimates should be sense checked with relevant care professionals, as the sample observed may not be representative of the entire activity, or the factors observed (eg consultation time) may not appropriately reflect the differences in total cost of different types of activity.

*Second option for Step 3: Using nationally collected or estimated information (eg reference costs) as weightings to allocate commissioner spend*

In some cases the relevant services may be covered by reference costs, which could be used as a source of activity data. In those instances local areas could also use the cost estimates submitted by the relevant provider if they are accurate (a case which is outside the scope of this handbook; see Figure 2). If that is not the case, local areas could use:

- average reference costs for a relevant benchmark of providers
- national average reference costs.

In other cases, the relevant services may not be covered by reference costs (or there may be concern about the accuracy of the reference cost information), but other estimates of average unit costs may be available.

In all three cases (average reference costs for a relevant benchmark, national average reference costs and other estimates of average unit costs), the cost information would be used as weightings to allocate commissioner spend to individual activity counts.

## Using average reference costs from benchmark providers

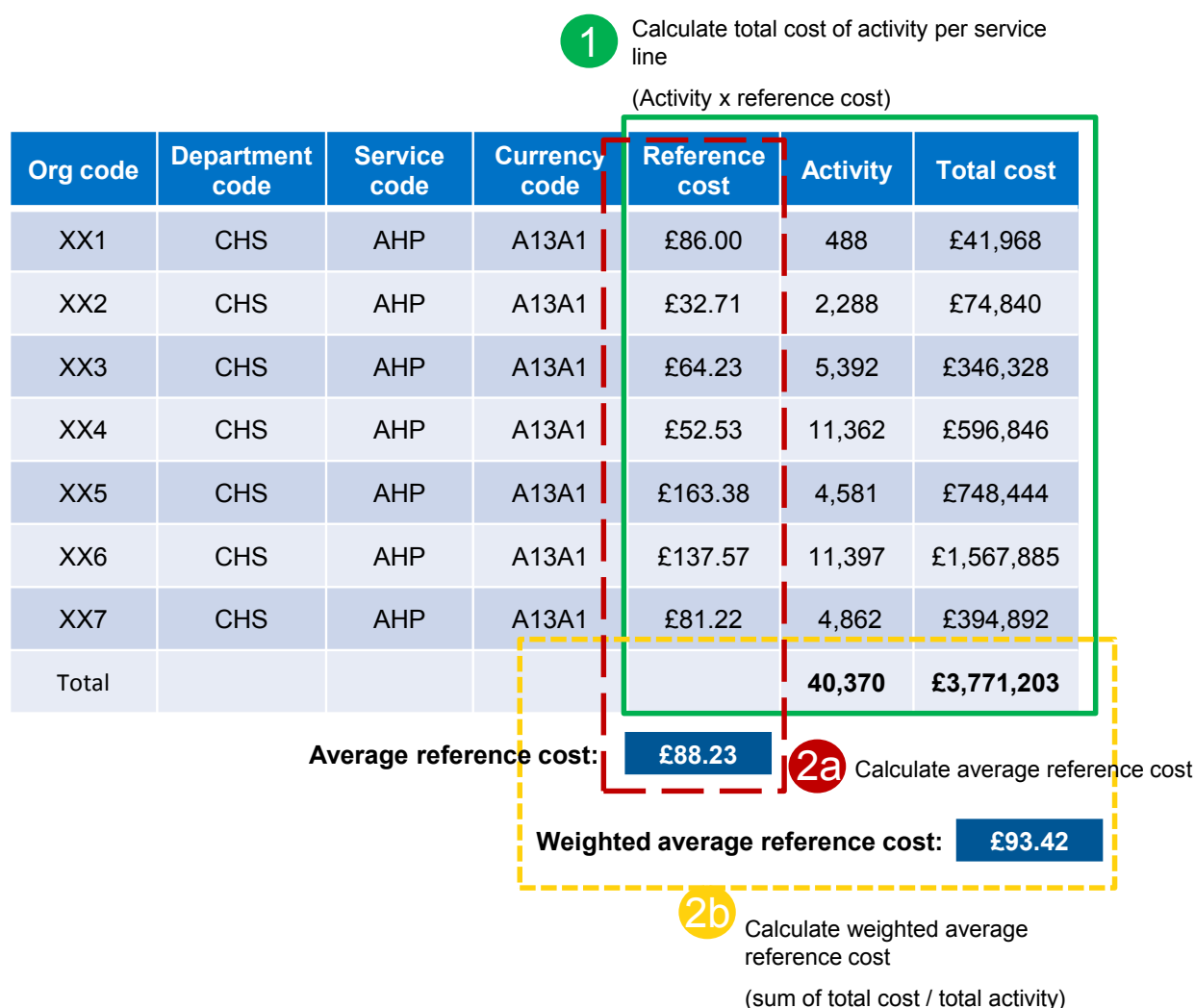
If local areas can identify other providers with a similar cost structure (and which serve a similar casemix), the reference costs from these providers could be used as a proxy for the costs of the relevant provider, and therefore as weightings to allocate commissioner spend to individual activity counts.

Figure 9 below presents an example of the calculations for a single type of activity. This process would need to be repeated for all relevant types of activity, and local areas would then use the resulting unit cost estimates as weightings to allocate commissioner spend to the different individual activity counts.

Local areas could choose either to calculate a 'normal' (unweighted) average or a weighted average. A weighted average would give more importance to providers with a larger volume of activity and may therefore be more representative of the true average across the entire benchmark. Local areas will need to decide which method to use, based on the composition of the benchmark, and how it relates to the relevant provider.

If the reference costs from the benchmark trusts are deemed reliable, this would provide a more accurate costing than using the national average. However, it does require the identification of a suitable benchmark and an assessment of the accuracy of the reference costs of the benchmark providers.

**Figure 9: Calculating average reference costs from benchmark providers – hypothetical example**



## Using national average reference costs

If the accuracy of the reference costs for the benchmark providers is uncertain, or deriving an appropriate benchmark proves too difficult, local areas may prefer to use the national average. Figure 10 below provides an example.

If the cost structure of the relevant provider differs significantly from that of the national average, or a large number of providers have submitted inaccurate reference costs, the national average may not be an accurate value to use.

**Figure 10: Using national average reference costs to allocate commissioner spend – hypothetical example**

**1** Identify relevant reference costs for services

**2** Calculate cost of activity based on reference costs.  
(Activity per service line x relevant reference cost)

Drug services (non-admitted)	Activity provided	Contract value	Reference cost category (NHS reference costs, alcohol services)	National average unit cost	Cost of activity (by reference cost)	Local cost weighting	Local average unit
Adult home visit	14,945		Adult, Community Contacts	£121	£1,807,108		£136.81
Adult clinic appointment	3,789		Adult, Outpatient Attendances	£60	£225,507		£67.34
Child & adolescent home visit	211		Children and Adolescents, Community Contacts	£273	£57,564		£308.67
Child & adolescent clinic visit	2,105		Children and Adolescents, Outpatient Attendances	£267	£561,383		£301.74
<b>Total</b>	<b>21,050</b>	<b>£3,000,000</b>			<b>£2,651,561</b>	<b>£1.13</b>	

**3** Calculate local cost weighting to match unit costs to contract service value.  
(Total Contract Service Value / Activity x Reference Cost Total)

**4** Apply local cost weighting to national reference costs to calculate Local Average Unit Cost.  
(National Average Unit Cost x Local Cost Weighting)

## Other available average unit cost estimates

If reference costs are not available for the relevant types of activity, or if there are concerns about their accuracy or applicability, local areas could turn to other available estimates of average unit costs (for use as weightings to allocate commissioner spend to different types of activity). For example, the unit cost estimates for a number of health and social care services published by the Personal Social Services Research Unit (PSSRU)<sup>36</sup> could be used if the approaches described above are not possible. But local areas should keep in mind that such estimates may not accurately reflect the circumstances of their local provider.

Last option for Step 3: no differentiation between types of activity (calculating average commissioner spend across total activity count)

Where neither of the above options can be used for this step because of a lack of available data, local areas will only be able to calculate an average commissioner spend per activity count for a service line (ie undifferentiated between the various types of activity so, for example, a home visit would be allocated the same £ value as a clinic consultation). This average spend can be calculated as the total commissioner spend on the relevant service divided by the activity count for that service. Figure 11 shows an example.

36 <http://www.pssru.ac.uk/project-pages/unit-costs/>

**Figure 11: Calculating average commissioner spend across total activity count – hypothetical example**

**1** Identify activity and contract value

Cases	Activity	Contract value	Unit cost
Phone consults	10,486		
Clinic visits	33,786		
Home visits	13,980		
	58,252	£2,500,000	£42.92

**2** Calculate average unit cost  
(Contract value / total activity)

This method is only accurate when the different types of activity have similar costs, as can be the case when a homogeneous service is delivered to patients (ie all patients receive exactly the same service). In the hypothetical example above, phone consultations, clinic visits and home visits would be allocated the same cost. Local areas using this method should be aware of the impact of this on the accuracy of the final estimates of commissioner spend per patient (or GP list). The accuracy will be even lower in instances where local areas have been unable to allocate the contract (£) amount to the different services (last option for Step 1).

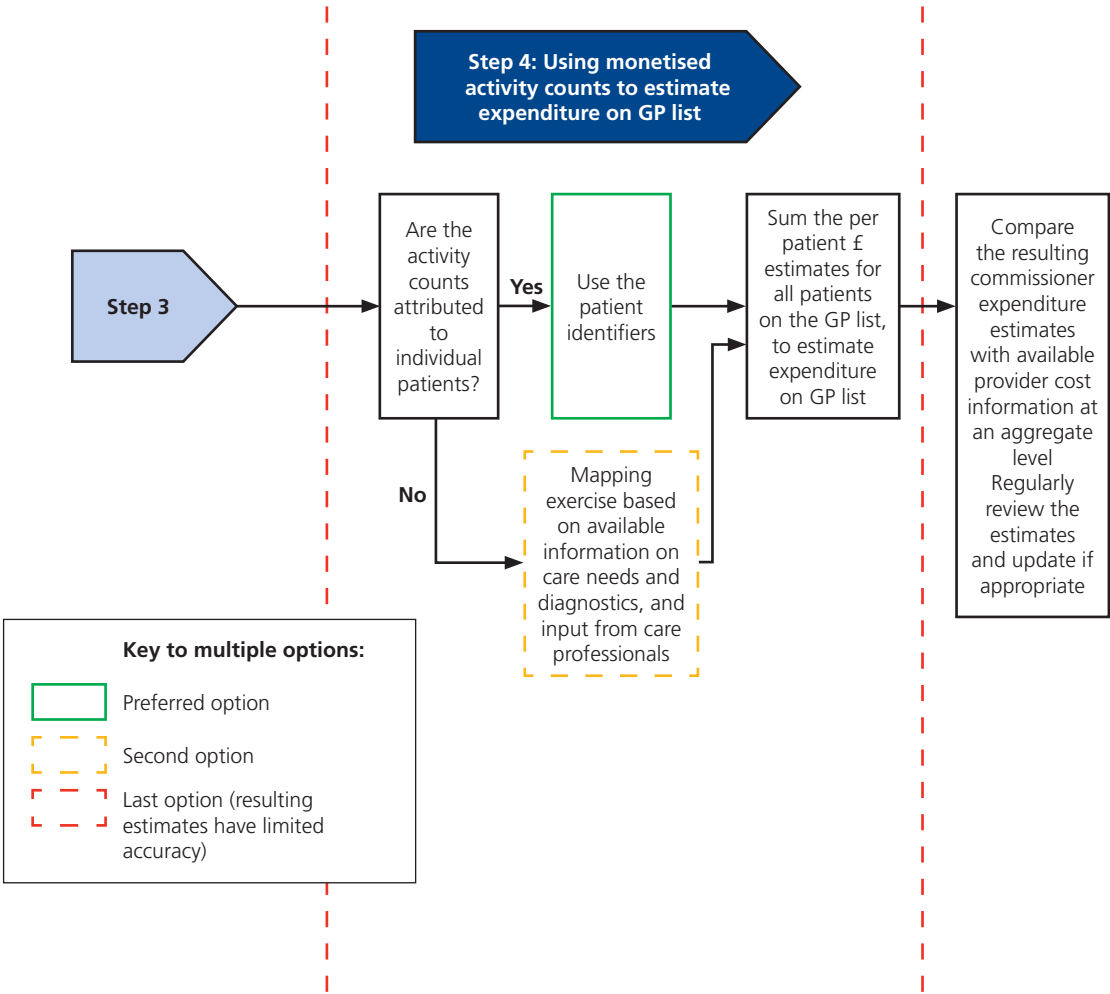
#### **2.4.5. Step 4: Mapping the monetised activity counts to individual patients and bundling them into GP lists (see Figure 12)**

On completion of the Steps 1 to 3 presented above, the contract amount (ie commissioner spend) should be allocated to the various activity counts collected.

Where activity counts have a patient identifier (eg NHS number or pseudonymised identifier), mapping these monetised counts to individual patients and rolling up the GP lists is a straightforward process.

Where activity information is not attributed to individual patients, a mapping exercise is necessary to allocate the (monetised) units of activity to specific groups of patients. The purpose is to try to identify the specific (groups of) patients which the relevant activity counts relate to, rather than allocating the spend on these services to the overall patient population.

Figure 12: Step 4 – Overview



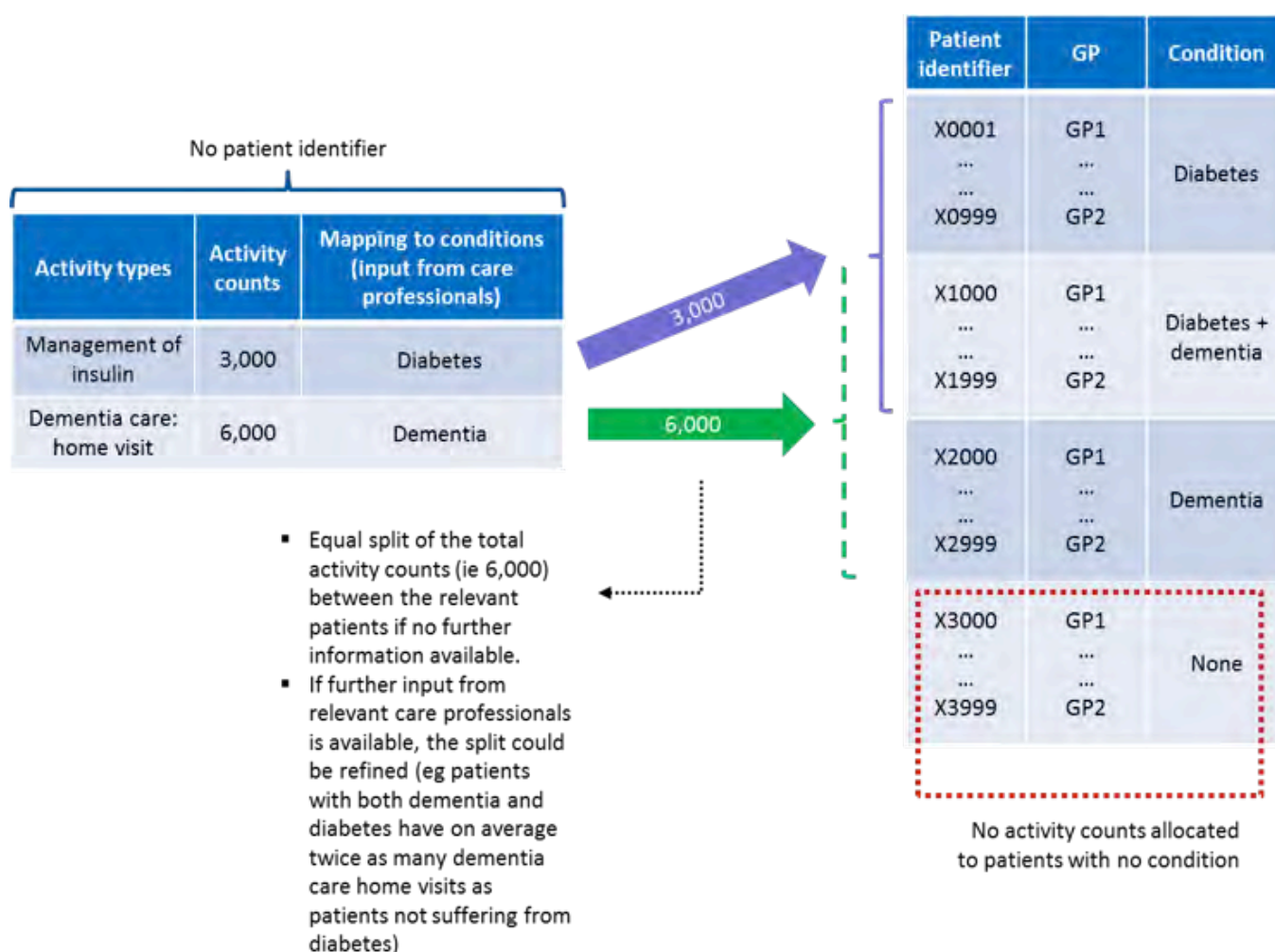
This is done by:

1. Local areas need to consider what information relating to care needs and diagnostics is available at a patient level, to group patients by needs and conditions (eg patients with a diagnosis of diabetes based on the quality and outcomes framework (QOF) registry).
2. Wherever possible each type of (monetised) activity should then be mapped to the group(s) of patients they are made available to (eg management of insulin services and the related spend could be allocated to patients with diabetes).

Clinicians and other staff familiar with the delivery of the relevant services should be involved in this process (for instance to provide knowledge about what specific types of care relate to specific conditions and about disease prevalence). In particular, local areas should seek input from the relevant GPs, who should have insight into their patients' needs across the full spectrum of care.

Figure 13 below presents an example of this process.

**Figure 13: Mapping the monetised activity counts to individual patients – hypothetical example**





The possible inaccuracy of the resulting spend estimates, especially at an individual patient level (eg two patients with diabetes may not receive the same type and/or amount of care) should be taken into consideration when using these estimates.<sup>37</sup>

Once commissioner spend has been allocated to individual patients, local areas should then bundle the patients into their relevant GP list(s), to estimate commissioner spend per GP list.

### **Case study: Encompass (Whitstable, Faversham, Canterbury, Ash and Sandwich) vanguard: Mental health service-level baseline contained in block contracts**

Encompass included a range of mental health services in its WPB, many of which were paid for via a block contract arrangement.

The vanguard included those mental health services it identified as being able to influence at a community level in its WPB: low and moderate acute mental health services, Improving Access to Psychological Therapies (IAPT services) services and services provided by the voluntary sector. The scope may change as the vanguard is further developed.

Although the acute mental health services were paid for via a block contract, these were split by service level in the contract. Mental health clustered activity was available at a GP practice level and could be used to apportion the expenditure. Acute assessments were not available at a practice level but this expenditure was apportioned based on activity usage of clusters.

The Encompass vanguard does not cover the whole Canterbury and Coastal CCG population, yet the block contracts for mental health did cover the entire CCG population. Therefore, activity and spend relating to out-of-scope GP lists had to be removed. To do this, a population proportional carve out was made using the weighted practice list sizes of the out-of-scope GP lists, as a proportion of the total GP list size covered by the block contract. This identified the mental health activity and spend for removal from the WPB.

## **2.5. Service-level spend on primary care**

The patient lists of participating primary care providers define the in-scope population of the organisation holding a WPB. Therefore, it should not be necessary to disaggregate primary care baselines by population cohort. However, it may be necessary to identify the monetary share of specific services if all primary care provision is not in scope.

The nature of primary care funding means there are a number of different funding streams available to primary care. Where the intention is to integrate primary care fully into the ACO, the primary care funding moving into the WPB will include as many of the existing funding streams as possible. However, there are three circumstances in which it may not be possible to move primary care funding streams into the WPB at the start of the contract period. These are:

- payments are activity-based and require incentivisation for public health reasons (eg vaccinations)
- funding is pooled centrally and individual practices apply to access this (eg GP Forward View (GPFV) funding)
- current legislative framework doesn't allow funding to be pooled within the ACO (eg dispensing doctors' payments).

37 See: <https://www.england.nhs.uk/wp-content/uploads/2016/07/mcp-care-model-frmrwk.pdf>



We are proposing that the majority of these payments flow through the fully-integrated ACO but are accessed according to the same criteria as under the current system. For example, payments for vaccinations and immunisation would be according to the activity performed.

The only exception to this is dispensing doctors' payments, for which the current legislative framework does not permit the ACO to access funding.

Where primary care activity data is available, some areas have chosen to apportion spend according to the different types of primary care activity, such as by appointment type (see the case study below). While this is not required to calculate the baseline, it does allow better insight into how spend on primary care could evolve as both volume of primary care activity and the share of non GP-led appointments changes.

The contract values for enhanced and other services should be included in the WPB baseline where they relate to in-scope services. The method for disaggregating block contracts (see Section 2.4.2) could be used if a contract covers more than one service, not all of which are in scope.

## 2.6. Current spend in social care contracts

As set out in *ACOs and the NHS commissioning system*,<sup>38</sup> current legislation governing the pooling of budgets across health bodies and between health and social care is complex. Commissioners must satisfy themselves that proposals to pool budgets in order to form a WPB are lawful, and that funds whose use is set out in law are not being used for an unlawful purpose. Annex B to the above document is intended to assist with this exercise.

Most social care contracts are either block contracts or care packages. Block contracts can be disaggregated for the in-scope services and population using the method described in Section 2.4.2.

### **Case study: Tower Hamlets Together vanguard: Value of a patient-level primary care dataset**

Tower Hamlets Together vanguard used a patient-level linked dataset to develop its WPB. This EMIS dataset contains patient-level details relating to appointment types (eg telephone, GP, nurse appointment, etc) and patient attendance: therefore, these appointments could be allocated to the correct patient cohort.

However, the costs of appointments at GP practices were only available at a total level. Therefore, the CSU split the attendances into GP appointments, nurse appointments, telephone appointments and other non-clinical/HCA appointments (after excluding any DNAs and other anomalies in the data), and determined the costs for each of these.

The commissioner spend for practice nurse consultations, telephone consultations and GP appointments across all condition bands was calculated as follows:

1. In 2013/14, primary care activity was commissioned partly by NHS England and partly by the local commissioners. Therefore to derive commissioner spend across primary care, the CSU needed to obtain the primary care budget for NHS England-commissioned services and CCG-funded locally enhanced service payments from FIMS returns.
2. This budget was then allocated to GP appointments, telephone consultations and practice nurse appointments based on the relative activity levels (%) for each within the year.

38 <https://www.england.nhs.uk/publication/acos-and-the-nhs-commissioning-system-document-6>

3. Adjustments were then made to reflect the lower price of delivering telephone appointments and practice nurse appointments (based on local judgements and discussion with the commissioner). These final percentage allocations were then applied to the available budget to allocate total cost values to GP appointments, telephone consultations and practice nurse appointments.
4. Finally, the total cost values for each appointment type were divided by the total activity levels for that appointment type, to calculate the underlying currency for that appointment type. This currency could then be allocated to the appropriate cohort of patients based on their activity levels.

Further work is planned to validate these initial assumptions by engaging with primary care stakeholders to check and challenge.

Social care packages combine the services that together meet a person's assessed needs under their care plan. The value of these care packages belonging to people on the participating GP lists contributes to the baseline. But if the scope of services in the care package exceeds that in the WPB, the total care package value will not contribute to the baseline. If the value of the care package is not split by service, the method for disaggregating block contracts could be used to calculate the social care baseline.

The information governance arrangements for sharing data between the CCG and local authority could be a barrier to implementation. The CCG and local authority should actively engage to resolve issues ahead of time.

### **Case study: Tower Hamlets Together vanguard: Calculating the service-level baseline of social care packages**

Tower Hamlets Together vanguard decided to include adult social care services which are delivered to individual patients (placements, domiciliary care, personal budgets and meals) in its WPB analysis. Services not attributed to individual patients, such as social workers, were not included in the analysis as these cost elements were not directly attributable to specific patients in the linked patient-level dataset.

To calculate the value of these individual social care packages the vanguard obtained full year costs, split by care package, from London Borough of Tower Hamlets Council (LBTHC). This data was contained in the borough's Framework-I system and included information on the condition for which social care support is being provided.

The costs of these care packages were not disaggregated based on the provider of the support. Instead, the vanguard used the total sum of the care provided by both private providers and internal borough provision. It could therefore allocate the total incurred costs of these care packages, regardless of provider, to the appropriate patient cohort when setting the budget baseline.

The CSU and vanguard faced a number of challenges in obtaining the relevant social care data: the need to obtain patient consent; the need to follow information governance procedures; and holding the data in a safe haven. Work is also ongoing to ensure that patient-level social care data is shared with the CSU in a timely manner.

The signing of data sharing agreements between the vanguard and LBTHC was essential to being able to include the value of social care packages in the baseline.

## 2.7. Convert the current spend to the WPB baseline

So that the WPB baseline reflects the value of in-scope services delivered in the first year of the contract, the commissioner spend calculated using the above steps and methods should be adjusted to account for current year efficiency assumptions and planned service investments.

Commissioners and providers may also locally agree other adjustments. The process of calculating current commissioner spend (such as by disaggregating historical block contracts) may identify instances where current spend on a service line does not reflect current activity, efficient provider costs or affordable commissioner spend. For example, community service providers may better support overall business viability by internally cross-subsidising between service values set out in block contracts, rather than seeking to adjust the service line values in the contract. Such cross-subsidies would need to be unwound when the service line moves to ACO scope from the current block contract.

The initial baseline value that is calculated should also be tested against historical data to identify volatility or one-off effects in the reference year. For example, the effects of changes to the service delivery model part way through the baseline calculation period may need to be adjusted for when calculating the WPB baseline.

These adjustments aim to ensure a realistic match between the contractual requirements and the WPB value. That match will be examined in the Integrated Support and Assurance Process (ISAP) before an ACO contract is awarded.

To avoid muting the effect of intended policy on provider efficiency, the current spend should not be adjusted to:

- match the costs of an inefficient provider or
- reduce the gap between national activity-based prices for secondary care (which determined the current year's acute services baseline) and reference costs.

## 3. Estimating WPB values for future years

### 3.1. Introduction

ACO care models are expected to hold long-term contracts, potentially running up to 10 years. While it will not be feasible to agree precise contract values for the full term of the contract, indicative Whole Population Budget (WPB) values will need to be provided for the full contract term as part of the procurement process. Initial procurement and contract documents will need to specify these annual estimated WPB values. This chapter describes the process for estimating the WPB values for future years.

Where providers and commissioners agree that certain circumstances could trigger changes to previously agreed budget values, this will need to be specified in initial documentation including: the triggers, mechanism for adjustment, frequency of potential adjustments and any tolerance thresholds for adjustments that can be made without triggering a re-procurement. Chapter 4 provides guidance on this.

Commissioner allocations act as an overall constraint on the level of funds available for health and care services in the local area. This together with the considerations set out in Section 3.6 must be recognised in estimating the WPB value for future years.

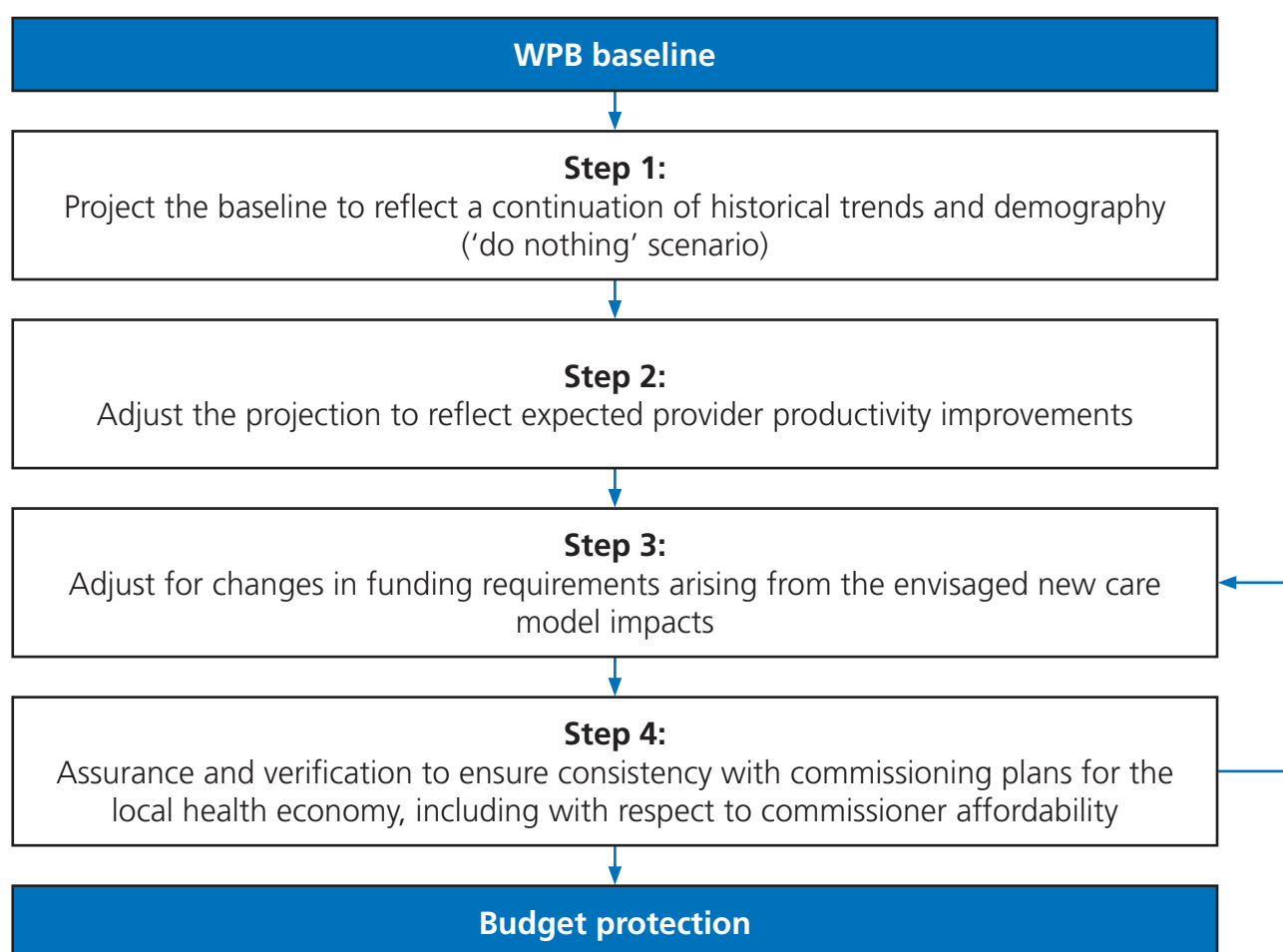
The scope of a WPB is likely to include services funded from multiple different streams, such as:

- core CCG funded services
- primary medical care funded services
- social care funded services
- public health funded services.

The future values of each funding stream will determine the constraint on the estimated overall WPB for future years and each commissioner's funding contribution to the WPB. This will need to be taken into account at the relevant steps in forecasting the WPB.

The steps in forecasting WPB values for the duration of the contract are shown in Figure 14 below. The starting point is the WPB baseline (described in Chapter 2).

**Figure 14: Steps in forecasting the WPB**



In some circumstances (see Section 3.2.1), growth in commissioner allocations can provide a good starting point for forecasting the WPB that is consistent with both commissioner affordability and service sustainability.

Any issues identified at Step 4 may require adjustments to the new care model and the commissioning underlying Step 3.

Where commissioners plan to modify the scope of the services or the population covered by the WPB in future years (eg to bring mental health services into scope in year 4), such changes need to be factored into the budget calculation for the relevant years and specified upfront in the contract. Wherever possible, these values should also be clear at the outset of procurement. For further information about the general principles for ACO procurements, see the *Procurement and assurance approach*.<sup>39</sup>

This chapter focuses mainly on the use of planning assumptions for healthcare services. Where other services are included in the ACO service scope, it may be possible to follow the approach outlined here, substituting the assumptions for healthcare with appropriate estimates for other types of services.

39 <https://www.england.nhs.uk/publication/procurement-and-assurance-approach-document-4>

## 3.2. Considerations for forecasting

### 3.2.1. Commissioner allocation growth rates

Under a few specific circumstances, allocation growth rates for CCG core funded services and primary medical care services may, as a composite measure of activity and cost pressures, be used as a starting point to project the WPB baseline forward.<sup>40</sup> This approach is appropriate where (a) the ACO population and service scope is significantly aligned with the population and service scope for relevant funding streams and (b) the cost of ACO services is expected to grow at the same rate as the average growth across all services within the commissioner(s)' allocation(s) for each stream.

Where the WPB scope includes services that range across a number of different funding streams, the budget forecast for those services should be aligned with growth in the relevant funding streams.

For contracts where the scope of the ACO is intended to increase over time to closely match the scope for a particular funding stream, local areas can move to using the rate of growth for the relevant stream to inform the WPB forecast, rather than more granular approaches.

For periods beyond those for which commissioner allocations are available, the WPB forecast should be calculated using other national and local information, and the sources and process clearly documented in the procurement and contract documentation.

### 3.2.2. Key factors affecting future WPB values

Where the ACO provider's population and service scope do not significantly match the commissioner(s)'s, the following factors which reflect the expected costs of service provision over time need to be taken into account to estimate the WPB in future years:

- **Demographic change:** the size and make-up of the population, which are fundamental drivers of future health care demand.
- **Non-demographic activity growth:** the tendency for activity to increase over time due to factors over and above demographic growth, for example due to technological development and medical advances.
- **Inflation:** the general tendency for health and care input costs to increase over time as set out in NHS Improvement's assumptions for annual increases in unit costs.
- **Efficiency:** each year providers are expected, and incentivised, to find new and improved ways to deliver care more cost-effectively. NHS Improvement sets out the national tariff efficiency factor for each year and has provided planning assumptions through to 2020/21.
- **New care model impacts:** the anticipated funding requirements associated with implementing the new care model and expected usage of ACO services over time, plus any efficiencies over and above those expected of the NHS more broadly as the new care model scales up.

In agreeing multiyear budgets to be held by vanguards, commissioners and providers will need to consider system sustainability, ensuring that the budgets are of a size that can be afforded, taking account of services within and outside the vanguard's scope and the need to manage risks.

<sup>40</sup> Many components of the CCG allocations are available at GP practice level so a target allocation could be constructed for a local area that does not match directly to the membership of the CCG. See the allocations technical guidance for further information: <https://www.england.nhs.uk/2016/04/allocations-tech-guide-16-17/>

### 3.3. Step 1: Projecting the WPB baseline

For local areas using commissioner allocations as a starting point for forecasting the WPB, the approach to take for Step 1 is described below (Section 3.3.1).

In other cases, the first step to projecting the WPB is to reflect anticipated future cost and activity pressures under business as usual:

- Local areas should use the national assumptions about expected future system pressures published by NHS Improvement and NHS England as a starting point to understand the scale of the local challenge to be addressed by the ACO. This is described in Section 3.3.2.
- Local information should then be used to refine these national assumptions. This may also be used for contract periods that extend beyond the period for which national economic forecasts are available. This is described in Section 3.3.3.

#### 3.3.1. Use of commissioner allocation growth rates

As set out in Section 3.2.1, commissioner allocation growth rates may, in some circumstances, provide a pragmatic starting point to inform the trajectory of the WPB.

Where multiple funding streams contribute to the WPB, commissioners will need to agree how they use the different forecasts in their calculation of the overall WPB for an ACO (eg by uplifting each budget separately and then combining them). The agreed approach should be recorded transparently.

Information about the drivers and method used to calculate growth rates for CCG core and primary medical care allocations is available in the *Technical guide to allocation formulae and pace of change*<sup>41</sup> published by NHS England. Because growth rates for each CCG are driven by both national and area-specific expectations, the rates in CCG allocations will differ by area.

Northumberland CCG is using its allocations growth rate to project its WPB baseline. The PACS model in the area is being designed to deliver a large majority of the services commissioned by the CCG, for its whole population.

The most recent CCG core and primary medical care<sup>42</sup> allocations and growth rates<sup>43</sup> have been set for the next five years (2016/17 to 2020/21), the first three years as firm allocations and the last two years as indicative. Any non-recurrent funding (eg General Practice Forward View funding) will not enter the WPB.

For other in-scope services, such as social care and public health, local areas choosing to use allocation growth rates as a starting point in determining the WPB forecast will be expected to use the relevant growth assumptions to forecast the relevant component(s) of the WPB.

41 <https://www.england.nhs.uk/2016/04/allocations-tech-guide-16-17/>

42 NHS England has published indicative allocations for other (non-medical) primary care allocations at CCG level. These mainly comprise community pharmacy, dental and ophthalmic services, all of which are outside the current scope of collaborative commissioning. These allocations are not currently robust enough at CCG level to be used in isolation and we strongly advise against sole reliance on them for CCG geographies.

43 <https://www.england.nhs.uk/resources/resources-for-ccgs/#finance>



### 3.3.2. National planning assumptions

National assumptions about future pressures on healthcare up to 2020/21 are available separately for:

- hospital and community healthcare services (HCHS)
- other healthcare services.

In many cases, the duration of contracts for ACO care models will exceed the period for which national economic indicators are readily available. In these circumstances, local commissioners will be expected to forecast economic indicators to inform their WPB estimates to be included in initial procurement and contracts.

While the cost pressures described in this section have been used to calculate commissioner allocations up to 2020/21 at a national level, allocation growth rates differ by CCG, as outlined in Section 3.3.1 above.

#### *Hospital and community healthcare services*

For HCHS, national assumptions are available that allow uncontrollable input cost and demographic activity pressures to be accounted for separately in baseline projections, as described below.

**Unit cost uplifts** Cost growth rates vary by type of healthcare input. NHS Improvement publishes assumptions on NHS provider inflation for different categories of inputs.<sup>44</sup> At the time of writing, assumptions are available for the period 2016/17 to 2020/21 (see Table 3 below).<sup>45</sup>

**Table 3: Assumptions on NHS provider inflation by input category**

Cost area	2016/17	2017/18	2018/19	2019/20	2020/21
Pay and pensions	3.3%	2.0%	1.6%	1.6%	2.9%
HCHS drugs	4.5%	4.6%	3.6%	4.1%	4.1%
Capital costs	3.1%	3.2%	3.2%	3.1%	3.1%
Other	1.7%	1.8%	2.1%	1.9%	2.0%

The economic assumptions in Table 3 do not include changes to bills for the Clinical Negligence Scheme for Trusts (CNST) and adjustments for new policies or service developments; the NHS Legal Authority assesses these annually. In 2017/18 and 2018/19 the anticipated increase in CNST contributions is 15%. In 2017/18 and 2018/19 the anticipated increase in CNST contributions is 15%.<sup>46</sup>

To apply these inflation factors to the WPB baseline, each needs to be multiplied by the proportion of costs it accounts for. The average weights across HCHS in the national planning assumptions for each input category are available via the link in footnote 44; the latest values are given in Table 4 below.

44 <https://www.gov.uk/government/publications/economic-assumptions-201617-to-202021/economic-assumptions-201617-to-202021>

45 These may be refined through the planning process for 2017/18.

46 <https://improvement.nhs.uk/resources/national-tariff-1719-consultation/>

**Table 4: Breakdown of unit cost projection weightings**

Cost area	Cost weight
Pay and pensions	65.44%
HCHS drugs	8.17%
Capital costs	4.76%
Other operating costs	20.17%
Litigation pressures	1.46%

Source:  
NHS Improvement, economic assumptions 2016/17 to 2020/21

Combining these inflation factors and cost weights generates the cost uplifts for HCHS applied to nationally calculated prices. An example of the calculation to generate the cost uplifts is shown in Table 5 and the assumptions for 2016/17 to 2020/21 are provided in Table 6.

**Table 5: Combining inflation factors and cost weights for summary cost uplift assumptions**

Category	Cost weight	2016/17 inflation factors	Cost uplifts
	a	b	c = a x b
Pay costs	65.44%	3.30%	2.16%
HCHS drugs	8.17%	4.50%	0.37%
Capital costs	4.76%	3.10%	0.15%
Other operating costs	20.17%	1.70%	0.34%
Litigation pressures	1.46%	17.00%	0.25%
Total			3.27%

**Table 6: Summary cost uplift assumptions**

	2016/17	2017/18	2018/19	2019/20	2020/21
HCHS cost pressure	3.27%	2.46%	2.17%	2.17%	3.04%

The example in Figure 15 illustrates the application of the cost uplift assumptions above to project unit costs for different types of acute activity.

**Figure 15: Using national assumptions to project unit costs for acute activity - hypothetical example**

HCHS cost pressures	16/17	17/18	18/19	19/20	20/21
	3.27%	2.46%	2.17%	2.17%	3.04%

Unit costs (£)	2015/16
A&E attendance	£125
Outpatient appointment	£110
Elective admission	£3,000
Non-elective admission	£1,500

Unit costs (£)	16/17	17/18	18/19	19/20	20/21
A&E attendance	129	132	135	138	142
Outpatient appointment	114	116	119	122	125
Elective admission	3,098	3,174	3,243	3,314	3,415
Non-elective admission	1,549	1,587	1,622	1,657	1,707

Activity pressures Healthcare activity increases over time due to demographic and non-demographic change. National assumptions for activity pressures by HCHS setting have been calculated by NHS England for the period from 2016/17 to 2020/21, as shown in Table 7 below.

**Table 7: Assumptions for activity projections by service area**

Service	2016/17	2017/18	2018/19	2019/20	2020/21
Acute (including ambulance)	See IHAM				
Mental health	1.9%	1.9%	1.9%	1.9%	1.8%
Community	3.3%	3.4%	3.4%	3.4%	3.3%

The Indicative Hospital Activity Model (IHAM) is an interactive tool, available on the Unify system or by email on request, which can be used by local areas to model projections for acute activity; it generates indicative activity figures based on historical trends and demography. The default growth assumptions in IHAM combine CCG-specific demographic pressures (these are different for each CCG and year) and historical national average non-demographic pressures. Users can overwrite the default assumptions with local information.

The example in Figure 16 illustrates the application of the assumptions from IHAM for acute activity growth to project activity volumes for different services.

**Figure 16: Using national assumptions to project acute activity volumes – hypothetical example**

Activity pressures	16/17	17/18	18/19	19/20	20/21
A&E attendances	2.28%	2.29%	2.33%	2.30%	2.26%
Outpatient appointments	3.77%	3.86%	3.87%	3.82%	3.76%
Elective admissions	1.96%	2.05%	2.06%	2.00%	1.95%
Non-elective admissions	2.07%	2.17%	2.23%	2.19%	2.15%

Activity	2015/16
A&E attendances	70,000
Outpatient appointments	230,000
Elective admissions	25,000
Non-elective admissions	25,000

Activity	16/17	17/18	18/19	19/20	20/21
A&E attendances	71,596	73,236	74,942	76,666	78,398
Outpatient appointments	238,671	247,884	257,477	267,312	277,363
Elective admissions	25,490	26,013	26,548	27,079	27,607
Non-elective admissions	25,518	26,071	26,653	27,236	27,822

Combining the projections for unit costs and activity provides total cost projections. Using the output tables from Figures 15 and 16, the example in Figure 17 illustrates how the total costs for different acute services are projected.

**Figure 17: Total cost projections for acute services - hypothetical example**

Unit costs (£)	16/17	17/18	18/19	19/20	20/21
A&E attendance	129	132	135	138	142
Outpatient appointment	114	116	119	122	125
Elective admission	3098	3174	3243	3314	3415
Non-elective admission	1549	1587	1622	1657	1707

X

Activity	16/17	17/18	18/19	19/20	20/21
A&E attendances	71,596	73,236	74,942	76,666	78,398
Outpatient appointments	238,671	247,884	257,477	267,312	277,363
Elective admissions	25,490	26,013	26,548	27,079	27,607
Non-elective admissions	25,518	26,071	26,653	27,236	27,822

Total costs (£'000)	16/17	17/18	18/19	19/20	20/21
A&E attendances	9,243	9,687	10,128	10,586	11,154
Outpatient appointments	27,113	28,852	30,620	32,480	34,726
Elective admissions	78,974	82,573	86,107	89,735	94,267
Non-elective admissions	39,530	41,380	43,222	45,128	47,499
Total	154,860	162,492	170,077	177,928	187,646

### Other healthcare services

Total cost assumptions for healthcare services other than HCBS have been developed by NHS England up to 2020/21 (see Table 8 below). These assumptions combine expected increases in both costs and activity and, as with HCBS above, indicate the pressures facing the system before planned efficiencies and initiatives are taken into account. The values in Table 8 should be applied to the baseline total costs for each service line.

**Table 8: Total cost assumptions for other healthcare services**

Service	2016/17	2017/18	2018/19	2019/20	2020/21
Continuing health care	6.5%	5.8%	5.5%	5.7%	6.6%
Primary care drugs	6.8%	6.9%	7.0%	7.2%	7.3%
CCG other primary care	6.1%	4.9%	5.3%	5.2%	5.2%
CCG other	6.1%	4.9%	5.3%	5.2%	5.2%
Primary medical care	5.6%	5.3%	5.3%	5.4%	6.4%

### Case study 1: All together better Dudley vanguard: Calculating the 'do nothing' scenario

All together better Dudley vanguard is using the stepwise approach described above to project the 'do nothing' funding requirements for its MCP model.

The starting point is the 2016/17 budget for each of the service lines the MCP plans to include in its scope. The vanguard has identified almost 3,000 separate budget lines. Within a service line, individual activities are assigned as either in or outside the MCP scope. This level of granularity provides the flexibility to develop and test the budget.

The vanguard is using Office of National Statistics (ONS) population projection estimates to uplift the baseline to reflect demographic growth. For both inflation and non-demographic growth, the vanguard has categorised current expenditure into 16 groups, allowing different rates of activity growth and cost inflation to be applied to each service line:

- for hospital services, the activity growth assumptions are based on the information available in IHAM
- for other categories such as community services and prescribing, the commissioner has used a range of local information, including area-specific historical growth rates and that drawn from the work developing the local STP.

The growth rates for the appropriate category are applied to the service lines sequentially, enabling the vanguard to identify the impact of each separately. The vanguard is undertaking an internal assurance process. This includes obtaining independent support to ensure that the growth assumptions used in the model have a robust evidence base.

Finally, the 'do nothing' projections are calculated as the sum of in-scope service line budgets for each year, against which anticipated changes from their new care models can be tested.

**Local authority commissioned services** For social care, in the absence of local information, we propose using the GDP deflator as the starting assumption to project spend. This assumption is consistent with the expectation that social care expenditure will be broadly flat in real terms over the period of the Spending Review settlement (ie to 2020/21). The GDP deflator is calculated and updated by the Office for Budget Responsibility (OBR).<sup>47</sup> Estimates at the time of writing are given in Table 9.

**Table 9: Total cost assumptions for social care expenditure**

Service	2016/17	2017/18	2018/19	2019/20	2020/21
Social care	1.5%	1.8%	2.1%	1.9%	2.0%

Source: Office for Budget Responsibility: Economic and fiscal outlook supplementary economy tables

**Public health services** The GDP deflator, as a measure of general inflation, may also be an appropriate starting assumption to forecast growth in public health services expenditure in the absence of alternative assumptions.

### 3.3.3. Refining national assumptions with local information

The national assumptions above provide a starting point for local areas to understand the external financial pressures facing services within the ACO up to 2020/21. Wherever possible, local information should be used to refine these assumptions to ensure the WPB reflects the local care model and the population and services it covers. Only drivers of costs and activity that are outside the control of the provider should be considered to ensure the estimated WPB values continue to place utilisation and efficiency risk with providers. Any local inflation and/or activity pressures over and above the nationally assumed levels will need to be supported by robust evidence to mitigate the risk of inflating the WPB over the life of the contract without clear justification.

The IHAM, described in Section 3.3.2 above, is one tool commissioners can use to help model the impact of local assumptions. It allows users to customise the growth assumptions for secondary care activity and recalculates activity trajectories accordingly. The model provides outputs for the following service areas: A&E attendances, outpatient attendances and elective and non-elective admissions.

Commissioners with access to linked datasets may be able to project activity and cost changes more accurately by taking account of their population's health status as well as age and gender (see the case study below). This method is likely to encourage clinical engagement in budget setting and greatly improve the transparency and accuracy of the adjustments to reflect the financial impact of new patterns of care (Step 3 below).

<sup>47</sup> <http://budgetresponsibility.org.uk/>



### **Case study 1: Tower Hamlets Together vanguard: Use of London-specific population projections**

Tower Hamlets Together vanguard is using an index based on the Greater London Authority (GLA) Capped SHLAA population projections to uplift their base year population, informing their WPB forecast.

For London-based systems, the GLA projections have relative merits over ONS projections. This is because the method takes into account future housing developments and migration flows into, out of and between London boroughs.

GLA population forecasts are available via the following link:

<http://data.london.gov.uk/dataset/gla-population-projections-custom-age-tables>

In using GLA projections as the basis for activity trajectories for the WPB, Tower Hamlets Together vanguard is using a different data source from the ONS population projections used to project CCG allocations. This difference in underlying data may be a factor in any misalignment between the MCP's budget trajectory and the funding available to the vanguard to be considered in Step 4.

### **Case study: Tower Hamlets Together vanguard: Projecting activity using a patient-level linked dataset**

Tower Hamlets Together vanguard is developing a whole population approach to shadow testing its MCP budget, to enable an assessment of different segmentation scenarios. Tower Hamlets Together vanguard has taken a 'bottom up' approach to developing data. Originally, it designed the budget for its MCP model around the 'integrated care cohort' of its population, defined using GP registered patient data and quality and outcomes framework (QOF) flags and other data sources (eg GP lists, social care and community information).

The baseline budget for the integrated care cohort was calculated by linking patient-level data, such as SUS and provider-generated aggregate commissioning reports, to patient NHS numbers to assign activity and unit costs to each cohort and service line. In addition to the cohort of adults using integrated care, Tower Hamlets Together vanguard is working to extend its model to include all adults and children, maintaining the granularity of linked data where possible.

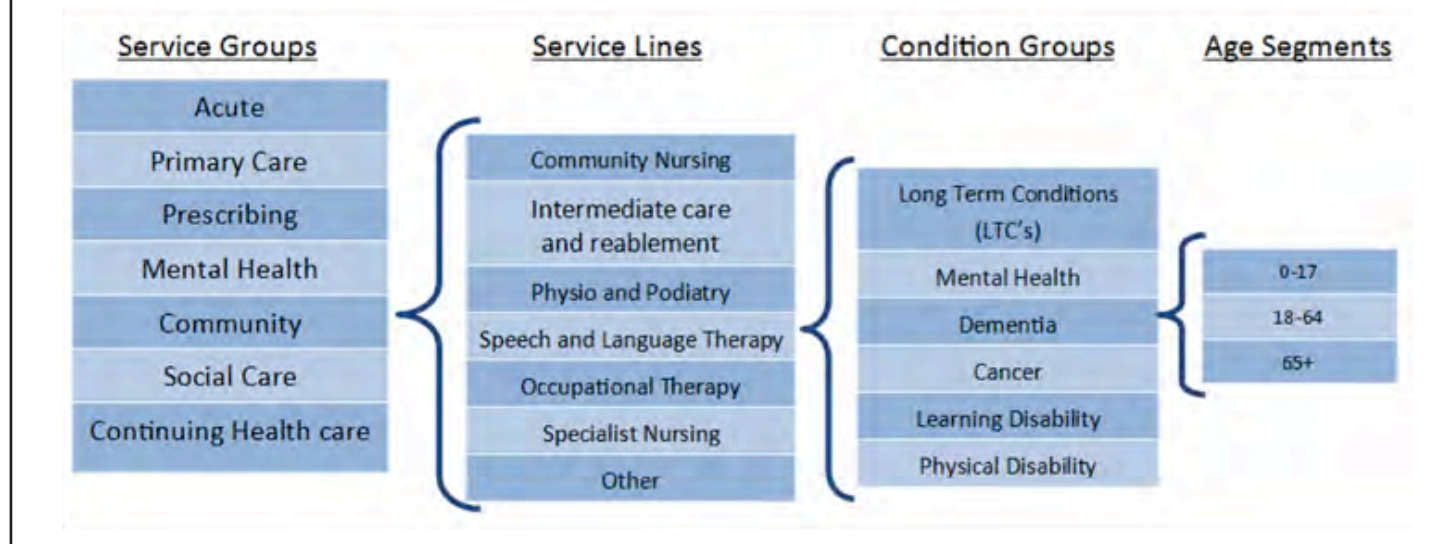
The intention is to project the baseline by considering population and demographic growth, unit cost projections and future service utilisation. The information in the patient-level dataset enables Tower Hamlets Together vanguard to develop its budget projection from information and assumptions about changing demand for different segments of the population:

Activity for each service is split by service line and condition, and then subdivided by age segment (as shown below). The average number of services used per patient is then calculated by using the patient NHS number to apportion activity to each segment. This base level of utilisation will then be projected using the assumptions defined above, with the resulting projections adjusted to reflect any expected reductions or shifts in activity as a result of the vanguard programme.

Multiplying the activity figures for each year by the projected unit costs determines the forecast level of expenditure.

#### **Service and population cohorts within the community service group**





### 3.4. Step 2: Adjust the baseline projection to reflect provider efficiencies

The expected multiyear WPB values agreed between providers and commissioners need to be adjusted to reflect provider efficiency requirements: acute providers are expected to achieve 2% efficiency improvements annually (see link in footnote 46). Efficiency targets for other services in the ACO should be agreed locally and reflected in the WPB. Because different services face different pressures and requirements, a single efficiency factor is unlikely to be applicable to the whole budget covering all services.

The funding for CCG allocations up to 2020/21 has been adjusted to reflect provider efficiency requirements. As such, where allocation growth rates have been used to project the WPB baseline in Step 1, no further adjustment is needed in this step.

### 3.5. Step 3: Adjust to reflect impacts of new care model

In addition to the provider efficiency improvement accounted for in Step 2, the new care models are expected to achieve further contributions to closing the gap between demand for services and available funding through commissioner allocations. These are likely to affect services both within and outside the scope of the ACO models, but may only be realised after an initial period of investment for service transformation.

Step 3 requires commissioners to decide how they should reflect the expected impact of new care models and other efficiency programmes<sup>48</sup> in the package of ACO financial incentives. Clinical and operational input is vital in this step to validate assumptions and provide the necessary detail of expected shifts in patient activity.

In particular, where local areas are planning to implement gain/loss sharing agreements (see Chapter 6), commissioners need to decide whether and to what extent the risks of realising these efficiencies are shared with providers through gain/loss agreements.

Local planning information, including vanguard value propositions, should be used as a starting point to understand the anticipated financial impact of the new care models. The IHAM tool described above can support commissioners at this stage: it allows users to investigate the impact of different programmes on secondary care activity trajectories up to 2020/21. The model is intended to encourage local and regional conversations and understanding about the effect of different assumptions on local circumstances. The current 'beta version' of IHAM does not include indicative numbers for efficiency programmes but a later release may do so.

<sup>48</sup> NHS England has identified six allocative efficiency programmes to achieve the NHS efficiency objective: new care models, right care, self care, prevention, urgent and emergency care and clinical prioritisation.

### **Case study: Mid Nottinghamshire Better Together PACS vanguard: Adjustments for planned efficiencies**

Mid Nottinghamshire Better Together PACS vanguard has reflected both national and local efficiency requirements in its budget forecast.

The starting point is to understand the level of planned efficiencies for each of the providers in the Mid Nottinghamshire Alliance and reflect this in the model. A similar process is undertaken in relation to system-wide transformation or quality, innovation, performance and productivity (QIPP) plans. Some QIPP plans are service specific while others are at a system level and will need to be distributed across services. This approach across providers and commissioners provides the total planned efficiencies for the system.

The alliance then allocates further efficiency improvement targets over and above those identified above. This could be based on the 2% tariff deflator if planned provider efficiencies do not meet this, or an expected efficiency where the system/provider benchmarks high for a particular service. The sum of planned efficiencies, transformation schemes and allocated efficiency improvement targets will form part of the projected WPB.

In addition, the Mid Nottinghamshire model calculates any residual unidentified and unallocated efficiency requirement. This is the difference between the WPB and the maximum affordable budget based on commissioner allocation and represents the challenge to the local health economy in future years. Responsibility for generating these savings can then be appropriately distributed across the system, based on the ability of different parties to respond. The Mid Nottinghamshire model therefore not only calculates the projected WPB but is also used as a planning tool to ensure that plans are developed and implemented to meet system-wide efficiency requirements.

The process outlined above is facilitated by the Mid Nottinghamshire model being structured as a commissioner–provider alliance rather than a prime provider or prime contractor model. By having the commissioner as a member of the alliance, responsibility for system-wide change is shared between providers and commissioners.

Areas that have well-developed local patient-level datasets may be able to model the impact of service redesign on the particular population segments targeted by the different initiatives.

For areas that used commissioner allocation growth rates in Step 1, the budget projections will still need to be adjusted for the financial impact of planned efficiency programmes and system changes. While efficiencies to be secured by local health economies are evenly divided across commissioner allocations at national level, the contributions of local programmes to delivering the national efficiency ask will not be drawn evenly from commissioning areas. As a result, adjustments are needed at local level for commissioners to ensure that the WPB projections reflect local circumstances and planning assumptions.

### **3.6. Step 4: Assurance and verification**

Steps 1 to 3 require commissioners to project the WPB baseline for their ACO provider to reflect future pressures on the system and to adjust this trajectory to reflect initiatives that will help to close the funding and efficiency gap.

The final step in the forecasting process is the formal assessment of the WPB forecast against overall commissioner allocations. This verifies that the estimated budget over the duration of the

contract is consistent with wider commissioning and service plans for the local health economy and local STPs.

Where multiple funding streams contribute to the WPB, the relevant allocation growth assumptions should inform the overall funding constraint on the local health economy and affordability of the WPB.

Any issues identified in this step would signal a need to reconsider the plans for the local health economy, and in the context of projecting a WPB, may lead to changes in Step 3 above.

From our work with PACS and MCP vanguards to date, we are aware that some local areas have addressed a mismatch between their funding allocation and the expected provider's budget requirement by including unidentified QIPP savings as an adjustment to the WPB projection. Where this approach has been taken, local areas will need to be assured that:

- the unidentified QIPP can be delivered by the ACO for the given population and service scope
- the ACO is the most appropriate place in the system to allocate this pressure, rather than it being given to or shared with services outside the organisation's scope.

### **3.7. Conclusion**

To forecast WPB values, commissioners are expected to understand what pressures will face the local system in future, particularly with respect to the ACO population and service scope, and to offset these against the potential impact of identified efficiency savings and initiatives. Commissioners then need to assure themselves that these plans are balanced, affordable within commissioner allocation constraints, and result in a WPB that reinforces the sustainability of the wider local health and care system.

The method used to develop budget projections should be transparently recorded, recognising that this information may need to be shared for future procurement activities and contract documentation.

## 4. Converting WPB estimated values into contractual values

### 4.1. Introduction

Chapter 3 outlines processes that commissioners can use to estimate WPB values for the duration of the contract. Commissioners will then need to convert these estimates into actual contracted WPB values and agree those values with the ACO provider. Commissioners can also use the provisions in the ACO Contract to periodically agree or update the WPB value in the context of the overall contract duration as long as the scale of change is allowable under the Public Contract Regulations (PCRs). As described in the *Procurement and assurance approach*,<sup>49</sup> commissioners will need to record the process they will follow to do this in sufficient detail in the procurement and contract documentation.

We expect the WPB will be agreed and updated periodically under two different processes:

- As part of a periodic agreement of updated or actual contracted WPB values in the context of the overall contract duration, so that the WPB can reflect relevant new evidence or information that becomes available during the life of the contract and that differs from that expected at the contract outset.
- Through a mechanistic update – as frequently as practical and at least annually – to accommodate patient choice, to reflect people moving in and out of the ACO area and to reflect the actual need profile of the population as new information becomes available, so that the WPB continues to operate as a capitated payment mechanism.

This chapter covers both of these processes, which will need to operate within the contract without necessitating a re-procurement. The guidance provided here should be used alongside the information in the *Procurement and assurance approach*.

In both cases, and as set out in Chapter 3, the commissioner allocations are an overall constraint on the level of funds available for health and care services that must be recognised. However, the WPB must not be adjusted mechanistically to reflect changes in commissioner allocations. Changes to commissioner allocations up or down may make it necessary to revisit the ACO contract and plans for the local area more broadly to ensure that delivery of services in scope of the ACO population is achievable within the WPB amount and that system sustainability is maintained.

The scope of a WPB is likely to include services funded from multiple different streams, such as:

- NHS core CCG funded services
- primary medical care funded services
- social care funded services
- public health funded services.

Once all relevant funding streams have been pooled within the WPB, the ACO provider will have the flexibility to determine the best use of these resources.

49 <https://www.england.nhs.uk/publication/procurement-and-assurance-approach-document-4>

## 4.2. Limits for agreed contract variations

Any factors affecting the WPB that commissioners and providers agree should be considered as an adjustment to WPB values estimated at the contract outset need to be within the following limits:

- the terms on which the contract was initially advertised (which would subsequently be reflected in the contract itself)
- the restrictions placed by the PCRs on in-contract variation.

## 4.3. Periodic agreement of actual contracted WPB values

While it will not be feasible to agree precise contract values for the full term of the contract, indicative WPB values will need to be provided for the full contract term as part of the procurement process. Even when the WPB values are fixed for each year of the contract at the outset, it will be prudent to agree a way to update the WPB in light of unexpected new information.

The duration of fixed WPB values and/or the frequency with which they should be updated will need to be determined locally. Local areas should make this decision taking account of pragmatic considerations about certainty over commissioner allocations and reliability of available forecasts. While longer agreed periods create greater incentives for the ACO provider to invest in care reform, improvement and population wellbeing, commissioners may not be in a position to carry the associated risk created by less certainty over future allocations and funding requirements across all services they commission.

It is important that adjustments are only made for materially significant differences, but not so large as to make it necessary to re-procure a contract or re-negotiate it beyond the changes agreed in variation schedules. Following this principle will help to ensure that WPBs create incentives for longer term planning by improving predictability and providing the stability to plan and implement changes as well as lowering administrative burden.

Factors that commissioners and providers may agree to take into account in the periodic agreement or update include:

- **Inflation:** The periodic agreement or update process set out in the contract could include an agreement to consider new and different national inflation forecasts notified by NHS England and NHS Improvement.
- **Impact of the ACO:** Where the impact of the new care model on cost efficiency or service utilisation is uncertain, the contract could include an agreement to periodically update the WPB value for the next fixed period based on observed efficiency and utilisation.
- **Other:** Any other variation which parties choose to agree to can be accommodated by the periodic agreement or update during the full contract period, so long as this is within the scope allowable under the PCRs. Further information is provided in the Procurement and assurance approach.<sup>50</sup>

The periodic agreement of contracted WPB values does not preclude mechanistic adjustments being made to reflect changes in the population covered by the ACO during these periods.

Where a contract value for the ACO for any period differs materially from the estimated WPB included in the initial procurement and contract documentation, it should prompt a review of other

50 <https://www.england.nhs.uk/publication/procurement-and-assurance-approach-document-4>

contract metrics as part of the update. For example, if the pace of service scope development differs substantially from that originally planned, the ability to manage risk, the profile of expected outcomes or other contract metrics may need to be adjusted. Commissioners must consider and document the flexibility of contract values in the original procurement process, and assess whether any material variations to original contract values would make it necessary to re-procure under current procurement regulations.

#### **4.3.1. Process for the periodic agreement of the contracted WPB**

**Step 1:** Commissioners and providers agree the factors that will be used to inform the periodic agreement of or update to the contracted WPB value. These factors are then written into contracts.

At this stage any thresholds for changes in the data that would prompt an update to the forecast values need to be agreed (Chapter 3).

Alternatively, local areas may agree to discuss whether an adjustment for inflation, for example, should be accommodated only if new inflation forecasts deviate from previous forecasts by a pre-agreed amount.

**Step 2:** As new information arises, or at previously agreed intervals, commissioners update the WPB calculation and assess the impact of the new information.

**Step 3:** At previously agreed intervals, commissioners and providers discuss the impact of the new WPB calculation and agree whether or not to update the WPB value from the original forecast and if so, by how much, according to criteria agreed in the contract.

As CCG allocations will be fixed within the period when these changes may apply, before agreeing an adjustment to the WPB, we expect commissioners will need to assess the impact of this decision on the funding available for other services, outside the scope of the WPB. In doing so, commissioners will be assessing the consistency of funding for services in scope of the WPB with those services out of scope.

#### **4.4. Mechanistic update for population changes**

As described in Chapter 1, WPBs are a simplified version of capitated payments. Capitation reimburses providers according to the size and needs of their population. The most sophisticated models include risk adjustments to account for the different needs of the individuals in the population.

The ACO population will change over time. While forecast and periodically agreed WPB values will reflect the projected changes in the size and make-up of the population, the agreed WPB values should be adjusted to reflect the (low value) impact of observed changes in the population. Commissioners and providers therefore need to agree a mechanism to adjust the WPB values, as frequently as practical, to reflect changes in the size and composition of the population.

This adjustment should be mechanistic to avoid frequent negotiation. That will ensure that the WPB operates as a capitated payment approach and supports patient choice, without undue burden from operating the payment approach.

This mechanistic adjustment ensures that the commissioner bears the financial risk associated with the size and composition of the population moving in or out of the ACO. This is consistent with the risk allocation that capitation aims to achieve, as set out in Section 1.1.2. It is important



that commissioners and providers are aware of the risks, and take local mitigating actions where appropriate. Some of these risks include:

- **Actual population size of the ACO may diverge from the agreed population levels.** Commissioners will need to fund any upwards variations using their contingency fund, or ensure that contracts with other providers mirror population movements for the ACO to facilitate movements across services.
- **Over (or under) payment for population changes due to insufficient data to calculate accurate financial values per patient type.** Commissioners and providers should agree the incremental impact of population movement on the mechanistic adjustment for WPBs as part of the contract value negotiations.

The most accurate approach to adjusting the WPB to reflect people moving in and out the ACO requires an agreed financial value per patient or per patient type. Many areas do not currently have patient-level linked datasets to provide a sufficient understanding of spend on or costs of different types of population segments to make such closely targeted adjustments.

During transition to mature organisations and until patient-level linked datasets are developed, we expect commissioners and providers to agree the value impact of people moving in and out of the population in procurement and contract documentation, and for the process to agree actual changes to the WPB to include some element of agreement between commissioners and providers. Simply using an average spend per head of existing population risks either significantly under or over reimbursing the provider depending on the type of individuals moving into/out of the population.

The frequency of this mechanistic adjustment will be limited by the available data and payment infrastructure. For example, currently primary care services funding is adjusted for every quarter, enabled by the frequency of updates to the registered lists.

We expect this area to develop over time and as ACOs mature, adjustments for people registering or ceasing to be registered with an ACO should become mechanistic so as not to require frequent negotiation.

NHS Improvement has published a tool that may help in estimating the impact of population movements<sup>51</sup> until local patient-level linked datasets are developed. Adjustments for changes in population health should only be made in some circumstances so that they neither discourage disease finding nor reward the provider for failing to improve its population's health.

## 4.5. Conclusion

Recognising the challenge in fixing WPB values for long contract terms, the ACO Contract will allow local setting of 'reset' periods. These allow the estimated WPB values to be updated to reflect relevant new information that becomes available and that differs from what was available at the contract outset.

WPB values may also need to be adjusted periodically within the reset periods to reflect changes to the ACO population.

Where agreed, the rules and mechanisms for adjusting the WPB need to be fully and transparently described in initial procurement and contract documents.

51 <https://www.gov.uk/government/publications/estimating-nhs-and-social-care-spend-a-tool-for-commissioners>



## 5. Improvement Payment Scheme

### 5.1. Introduction

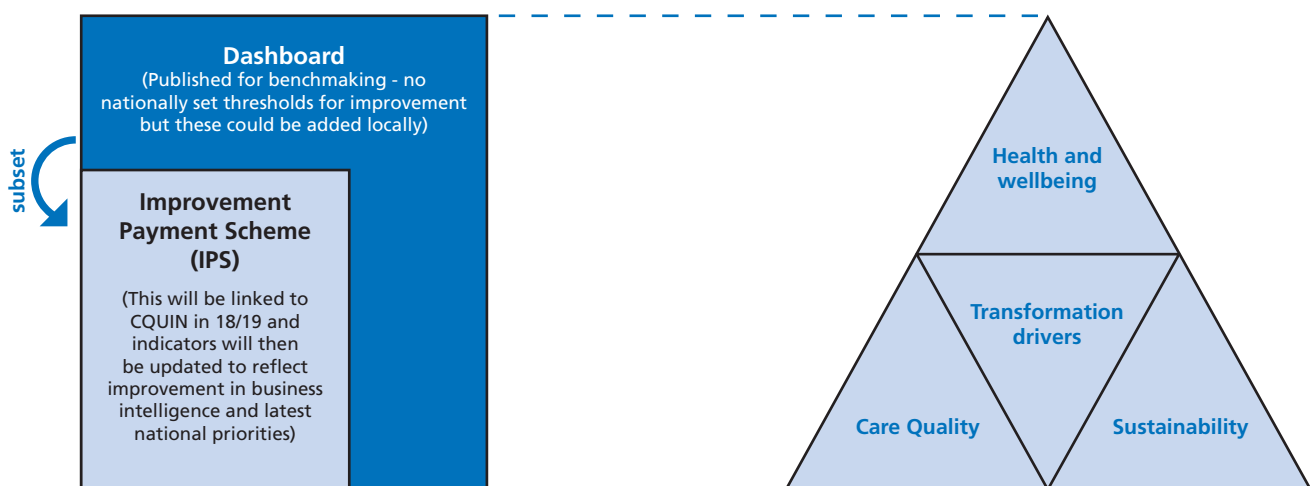
The purpose of the Improvement Payment Scheme (IPS) is to use payment to incentivise performance improvement across a small number of priority areas. Whilst the scheme will initially be aligned to the existing CQUIN scheme, and so largely process based, the metrics are intended to evolve to become more sophisticated and outcomes based over time.

### 5.2. ACO IPS

The IPS forms part of the incentives framework for ACOs as shown in Figure 18 below. The framework will provide an overview of:

- ACO performance
- contribution the ACO is making to the wider health economy.

**Figure 18: Overview of the incentives framework for ACOs**



The incentives framework for ACOs will be published and will include the IPS indicators. This supports our transparency drive to ensure that patients, the public and other interested parties can access this information. The appendix to the ACO Contract provides further information about the framework, including the proposed indicators.

### **5.3. Scheme quantum**

A proportion of the contracted WPB value is withheld to fund the IPS scheme. The amount that is paid depends on how well the ACO delivers the requirements set out in the IPS scheme. The quantum for the national element of the scheme will be designed to replicate the balance of financial risk and incentives that exist in the current national performance pay schemes. The implication of this is that the national IPS will be worth around 2.5% of a partially integrated ACO's contract value.

### **5.4. Commissioner flexibilities**

Sites involved in the co-design of this framework have advised us that they would value flexibility to reflect local priorities and ambitions. Commissioners can flex the IPS to meet their needs by increasing the quantum assigned to the IPS by paying against additional indicators.

The additional financial risk borne by sites choosing to include local metrics and assign additional contract value to them will be assessed through the Integrated Support and Assurance Process (ISAP) run by NHS England and NHS Improvement. Please refer to further guidance on the ISAP.<sup>52</sup>

### **5.5. Further information**

The incentives framework for ACOs document provides further information about the framework, including the proposed indicators.

<sup>52</sup> <https://www.england.nhs.uk/resources/resources-for-ccgs/#isap>

## 6. Gain/loss sharing

### 6.1. Introduction

This chapter explains the concept of gain/loss sharing, describing the important role it can play in enabling service transformation and providing guidance on the main design parameters. It provides a common structure, supported by illustrative examples and case studies, for commissioners and providers to use when locally introducing gain/loss sharing arrangements that facilitate delivery of ACO care models.

Gain/loss sharing was introduced as part of the local payment example (LPE) suite of documents published jointly by NHS England and NHS Improvement<sup>53</sup> to support the development of the new care models and the payment approaches needed to underpin them. This chapter builds on the LPEs by reflecting the early experience of MCP and PACS vanguards. It focuses on the design of the mechanism.

We are continuing to work with leading sites on a more detailed gain/loss framework which will be made available to support implementation. The national framework for gain/loss sharing will:

- describe indicators that ACOs should consider including in their local gain/loss mechanism; for example, around non-elective admissions, outpatients, elective care and delayed discharges. These will depend on the ACO's scope
- provide a suggested framework for establishing baselines, thresholds for payment and how risk is borne by the ACO and other parties
- provide a detailed step-by-step guide to planning, developing and implementing a gain/loss mechanism locally, to provide a level of national consistency and assurance
- highlight any practical lessons and case studies we have gathered from early work with vanguards
- offer any tools or workshop models that could be utilised by the sector.

An overview of the steps for implementing gain/loss sharing to be included in the national framework is provided in Section 6.4 of this handbook.

### 6.2. Definition

Gain/loss sharing in the context of ACO care models and the WPB is the sharing of savings (gains) or overspends (losses), generated through lower/higher than expected utilisation of a service, between one or more commissioners and one or more providers. For gains/losses to be transparently measured, they must be linked to specified activities; either in services provided by the ACO and/or in other services that are impacted by the ACO.

Gains and losses are calculated as the difference between the expected (ie baseline) commissioner spend on contracting for specified services for a defined population and the outturn actual spend, where outturn spend is determined by multiplying activity by a unit price. For example, if an MCP provides care closer to home (where clinically appropriate), there may be a larger fall in non-elective admissions in the acute setting than assumed in the baseline. If these admissions are reimbursed

53 <https://www.gov.uk/government/publications/local-payment-example-multilateral-gainloss-sharing>

through payment-by-activity, the commissioner spend will be lower than the baseline level, generating gains to be shared.

The gain/loss sharing mechanism is unsuitable for managing risks that sit solely with one party, for example risks associated with delivering cost efficiency. Building the mechanism on specified activities multiplied by prices helps to avoid any unintended risk sharing (eg cost efficiency risks).

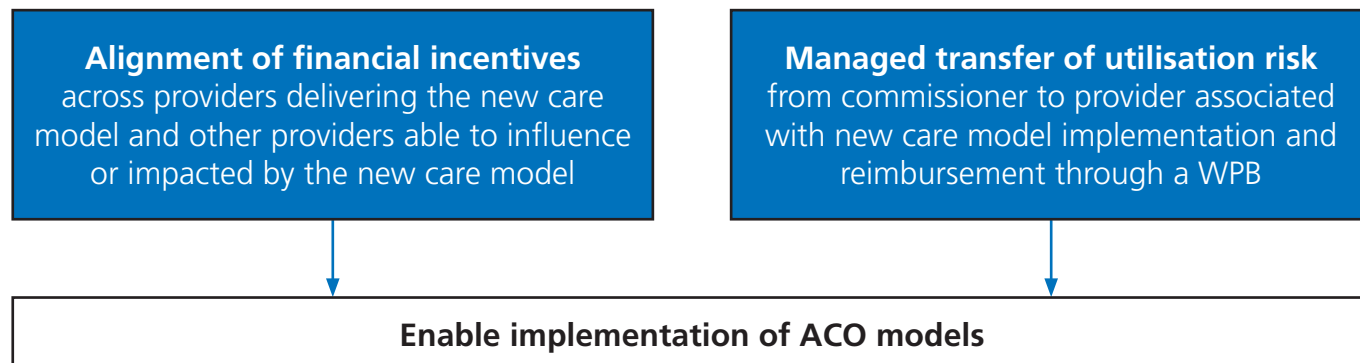
### 6.3. Context and purpose

The introduction of ACO care models involves significant uncertainty and associated risks around the impacts that these models will have on the use of services across care settings and providers, ie utilisation risk. How the ACO contract distributes these risks over time between commissioners and providers may help or hinder the successful delivery of the new care model.

The ACO care models also require much greater collaboration across current provider boundaries, possibly outside the ACO commissioned scope. This necessitates alignment of financial incentives. Otherwise it can be very difficult for providers to work together to deliver outcomes for the system as a whole.

Gain/loss sharing aims to support necessary service transformation through the two main channels shown in Figure 19 below.

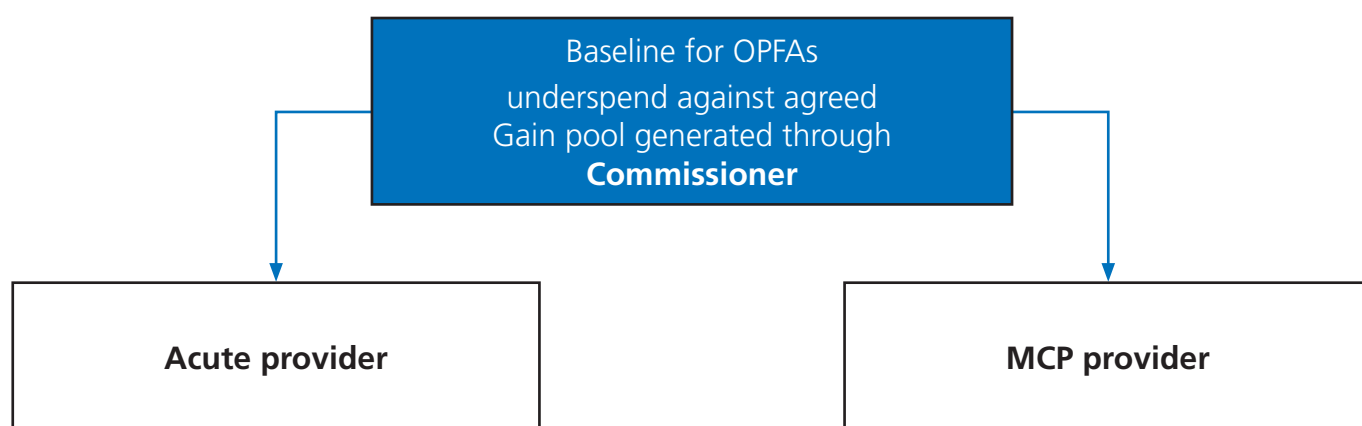
**Figure 19: Supporting service transformation with gain/loss sharing**



#### 6.3.1. Aligning financial incentives

Gain/loss sharing can give a number of parties a stake in the new care model, helping to align financial incentives. This is particularly the case if the arrangement involves all the parties that can influence and/or are impacted by the identified utilisation risks, including services and providers outside the scope of the WPB. This is demonstrated in the illustrative MCP example in Figure 20 below. The same rationale applies to other ACO care models.

**Figure 20: Illustrative MCP example of aligning financial incentives**



For simplicity, this multilateral gain/loss sharing example relates to a single service, outpatient first attendances (OPFAs), in the first year of the MCP Contract. The arrangement is one-sided so that only gains are shared; which in this example are gains made by the commissioner. Gains are generated through successful MCP care delivery resulting in a lower number of OPFAs by the MCP population and corresponding fall in baseline commissioner spend. The gains are measured using the relevant OPFA treatment function codes (TFCs) and their national prices, with patient NHS numbers or GP registration codes identifying use of OPFAs by the MCP population. The commissioner's contract with each of the providers specifies upfront that gains against the agreed OPFA baseline for the MCP population will be shared between the commissioner, MCP provider and acute provider, respectively. This helps align financial incentives across the system as follows:

- **MCP provider:** By sharing in any gains associated with OPFAs, the MCP has an added incentive to proactively reduce OPFAs compared to the baseline.
- **Acute provider:** The acute provider's share in any gains associated with OPFAs could help mitigate the impact of reduced revenue relative to their fixed costs and enable the provider to better engage in the overall service change.
- **Commissioner:** The commissioner, having oversight of the wider health economy, may retain a small proportion of the savings to make strategic investments in improved care delivery. The ability to retain some of the savings will also help to ensure commissioner engagement in whole system outcomes.

### 6.3.2. Managing the transfer and impact of utilisation risk

Payment for integrated care services via a 'fixed' WPB would transfer all utilisation risk (both upside and downside) from commissioner to provider. The level of risk placed on the provider could be substantial and unsustainable, especially where the provider is currently paid based on the actual activity it undertakes. Gain/loss sharing can help to manage the impact and transfer of utilisation risk over a number of years by allowing the ACO provider to mature during the contract.

## 6.4. Implementation framework

This section sets out a general framework – consisting of six steps – for commissioners and providers to follow when locally developing, agreeing and implementing gain/loss sharing

arrangements. It provides an overview of the key decisions which need to be made at each stage of gain/loss sharing and covers the design parameters which should be considered. This has been informed by our work to date with vanguards and international evidence.

The experience of vanguards suggests that local areas should begin gain/loss sharing discussions as soon as possible alongside their work to establish a WPB; given that arrangements may take time to design and be agreed. To aid effective discussions on gain/loss sharing and to avoid unnecessary duplication of tasks, before work starts on this framework we believe it's important for areas to have considered the key enablers set out in Section 1.1.5 and established a number of prerequisites. For example, it's crucial that the case for service change has been agreed, the proposed new care model has been designed and (as far as possible) the population scope and service scope for the new care model have been clearly specified.

Providers and commissioners should tailor arrangements to meet specific local objectives and/or to reflect local circumstances. For example, some vanguards have found it useful to establish a number of overarching principles and objectives to guide the process of agreeing a gain/loss sharing arrangement.

### Case study: Tower Hamlets Together MCP vanguard: Principles for gain/loss sharing

Below are Tower Hamlet's draft principles for developing a gain/loss share component as part of a population-based payment arrangement.

#### Gain and loss account

**Principle 1:** Gain loss sharing should facilitate a collective system transformation - working together to promote collaboration, transparency and change a culture of silo working, in order to deliver quality outcomes and what is right for the local population

#### Gateway review of gain and loss

**Principle 2:** Gain and loss sharing metrics should be measurable

**Principle 3:** There is agreement that there is a set amount of money in the system through the capitated budget ("one pound to go around"), therefore material gains/losses under collective responsibility need to be managed effectively in order to keep within this budget

#### Assessment of gain and loss

**Principle 4:** Gain and loss sharing rules should consider the evolution of the programme (i.e. gain sharing rules should change over the period of the contract)

**Principle 5:** Each organisation should be held accountable for contributing to delivery of required outcomes and metrics; risk-sharing should not dilute accountability of individual members for their own risk mitigations and delivery decisions. Where a collective responsibility is identified, the sharing of risk and reward will only apply to organisations whose actions directly impact on the desired result or change required

#### Model results in gain and loss share

**Principle 6:** Financial distribution of gains and losses should consider the *net* outcome across the system, rather than outcomes from individual service lines within the contract

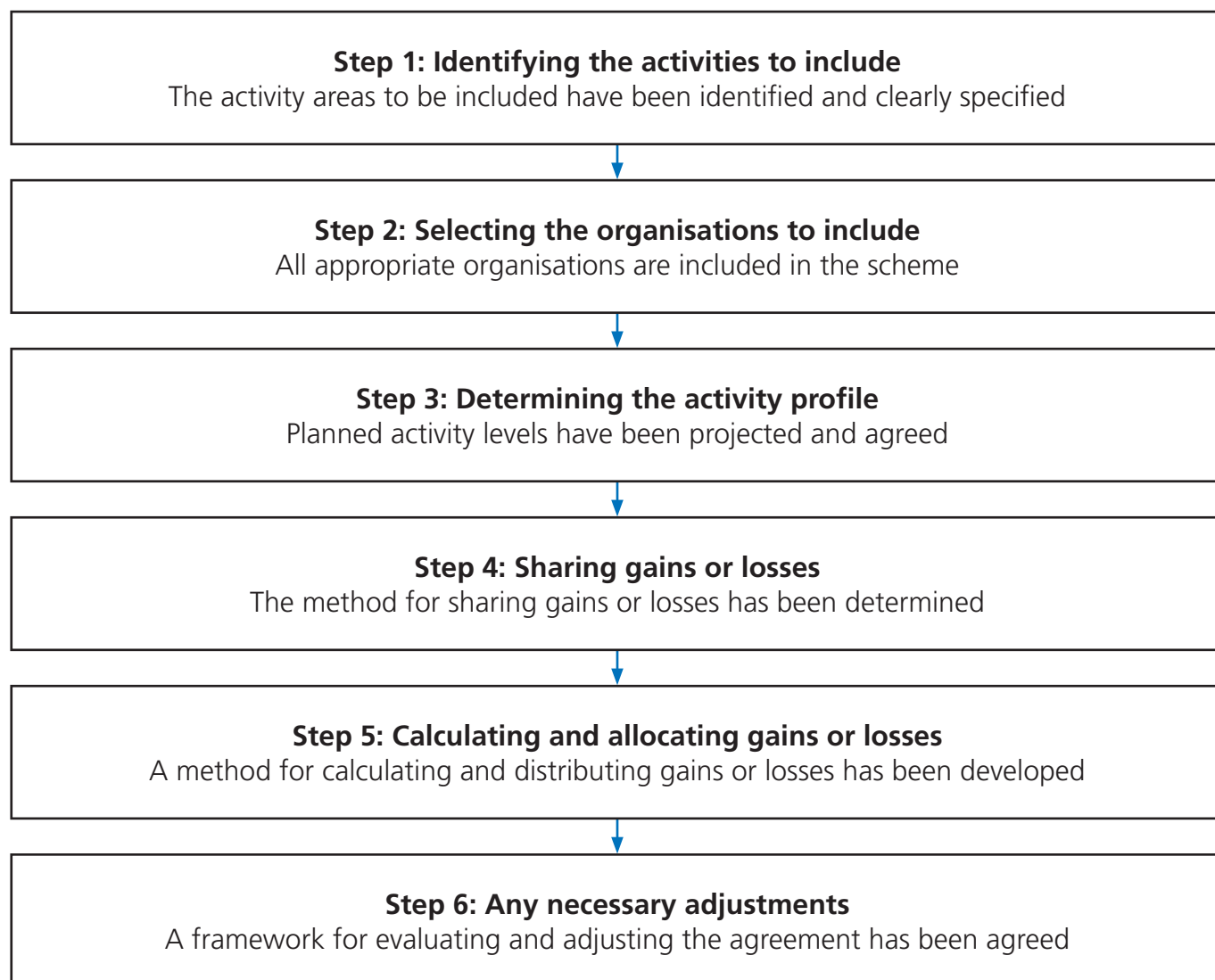
**Principle 7:** Where financial performance exceeds expectations, a proportion of these gains should be distributed to transformation funding / investments that will drive longer-term value for the system

In addition to gain/loss share, achievement of agreed outcomes will be part of the payment arrangement. The outcomes payment will be a key lever to ensure care delivers the outcomes that people want and need.

### 6.4.1. Gain/loss sharing implementation framework

Figure 21 gives an overview of the suggested framework. This has been written as a series of sequential steps. However, local areas may want to consider different steps at different times and to take some decisions, such as the risks and organisations to include, in parallel. Furthermore, the framework is by no means comprehensive, and each step will need to be explored and interpreted further by localities, applying their own circumstances as appropriate.

**Figure 21: Implementation framework**





### 6.4.2. Step 1: Identifying the activities to include

Local areas must determine and clearly specify which service activity or activities, and the associated utilisation risk, form the basis of their gain/loss sharing arrangement. The identified activity included in each arrangement will vary from one LHE to another according to local objectives and priorities. The principles underpinning the proposal and selection of an activity area should be consistent. Broadly speaking we believe these principles can be divided into a number of design and practical considerations:

- **aligned to new care model aims:** service activity or activities identified in value propositions, logic models and/or other relevant proposals as areas a new care model will influence
- **high degree of uncertainty in levels of demand:** such that gain/ loss sharing can help to allocate the financial risks to both commissioner(s) and provider(s) more fairly
- **multiple parties in the LHE are able to influence or are impacted by the service area**
- **material financial impact:** the identified system risks are significant
- **sufficient data is available** for accurate measurement, reporting and financial transactions
- **the incentives created by applying gain/loss sharing will not conflict with those created by other elements of the funding system** (eg IPS)
- **ability to minimise any perverse incentives which may be created**
- **all relevant parties can be held to account.**

Service areas that satisfy these criteria are most suited to gain/loss sharing agreements for the distribution of utilisation risk across the LHE. As well as the criteria above, local areas should also consider the level of influence the new care model will have on activity areas. For example, it might not be best to include all the activity associated with a specific service when the care model can only affect a subset of that activity.

#### Key lessons and principles

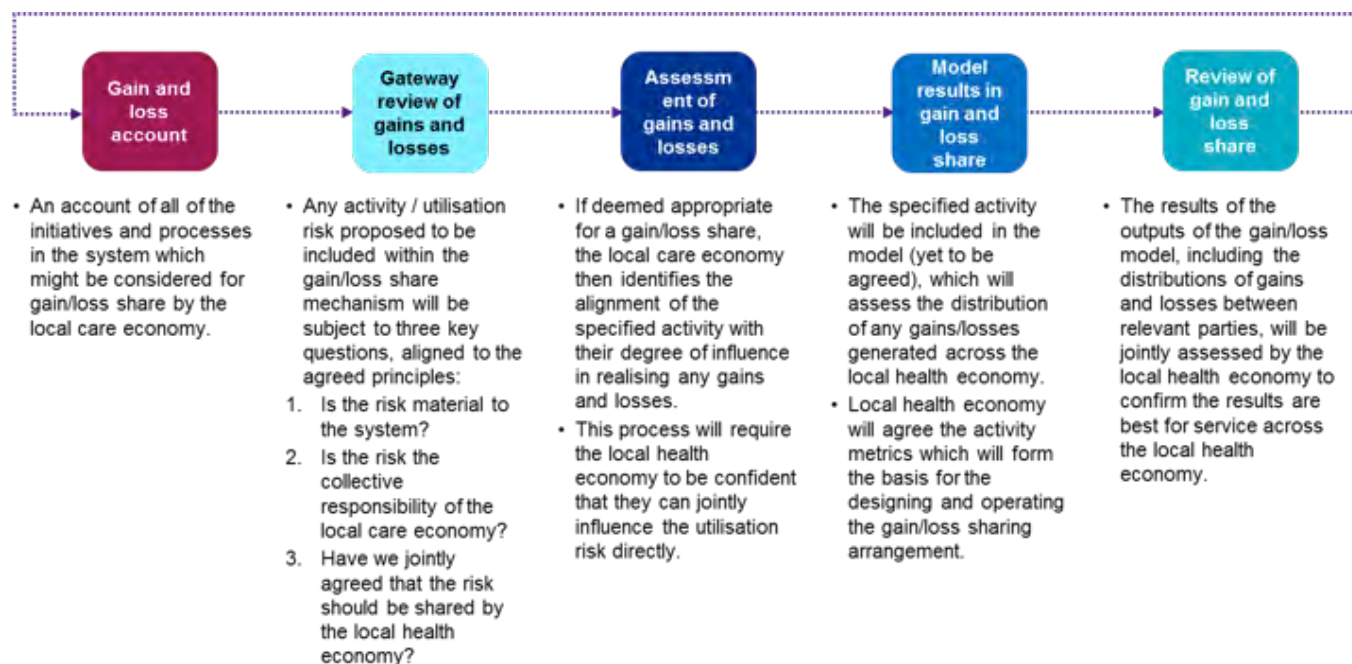
Both commissioners and providers, together with input from clinicians as appropriate, should determine which service(s) and associated utilisation risk(s) should be the shared responsibility of all parties.

Gain/loss arrangements should focus on risk sharing and not risk transfer. The mechanism should not shift risk around the system. It should better manage specific risks by sharing them more widely across the system.

Gain/loss sharing should not dilute the accountability of individual parties for their own risk mitigations and delivery of quality services for patients.

## Case study: Mid Nottinghamshire Better Together PACS vanguard: Decision-making process for gain/loss sharing

The Mid Nottinghamshire Better Together PACS vanguard, in discussing which gains/losses to share, is focusing on risks that are significant in size and which are deemed to be the collective responsibility of the group, such that the group as a whole should be accountable for those risks. The decision-making process that the site is using to establish whether a particular risk may be appropriate and beneficial to share is shown below.



### 6.4.3. Step 2: Selecting the organisations to include

Selection of the parties to include in the agreement will be closely linked to the activity areas that are to be covered. Local areas should start by 'identifying the players': all those organisations which can influence the service and/or activity identified for the agreement.

Prime consideration should be given to those organisations with the greatest ability to bear risk and their prominence in the scheme. For example, if gains will take the form of reduced expenditure on activity paid for on the national tariff, then the CCG will need to play a key role in any agreement to share these gains around the local health economy. It might be appropriate to exclude from any arrangement those organisations that can't bear risk or that have limited influence over the targeted activity.

### Key lessons and principles

Both commissioners and providers should participate in gain/loss sharing development from the outset if possible, to ensure co-development and to reduce the need for rework later down the line. A clear process should be established for developing gain/loss sharing that is jointly agreed by commissioners and providers.

Continuity in attendance between workshops is recommended as it limits the reopening of issues and enables progress to be made with each additional workshop.

#### 6.4.4. Step 3: Determining the activity profile

Planned activity levels for the covered service(s) will need to be projected for the length of the agreement (though they may be periodically adjusted – see Step 6 in Section 6.4.7). Where a service is included in a WPB the activity projection should mirror that in the forecasted WPB baseline. This will help with consistency and to avoid any double counting in the system. Where the service is outside the scope of the WPB (eg acute services), activity should be projected using a combination of historical trends and population and demographic projections, and the expected impact of the new care model. Any gains or losses will be calculated by comparing actual outturn activity against these planned levels (as per Section 6.2).

The likelihood of any gains or losses being generated will depend heavily on the forecasted baseline. Overall, we believe a two-sided agreement should aim to set the baseline so that the possibility of making a gain or a loss is roughly equal, to ensure incentives remain balanced. Where this is not possible or appropriate, local areas may want to adjust some of the design characteristics in Step 4.

#### 6.4.5. Step 4: Sharing gains or losses

A number of factors influence how any gains or losses should be shared between commissioner(s) and provider(s). For example:

- **The likelihood of generating gains/losses:** How gains/losses are shared between commissioner(s) and provider(s) should take into account the likelihood of them being generated. This greatly depends on how the baseline has been set.
- **Ability to influence risk:** Gain/loss shares should be aligned to each party's ability to influence utilisation risks associated with specific services and/or according to how parties are impacted by those risks.
- **Ability and appetite to bear risk:** Appropriate risk shares between parties will also depend on their willingness to bear risk.
- **Risk exposure over time:** The level of risk exposure (both downside (losses) and upside (gains)), will likely need to change over the duration of the arrangement as the new care model becomes established and organisations mature.
- **Maximum risk exposure:** The maximum amount of risk each organisation can bear from gain/loss sharing will need to be considered alongside any other risks, eg through the Improvement Payment Scheme (IPS).
- **Minimum savings rate:** It may be appropriate to establish a minimum savings rate, with only gains/losses over and above a minimum level being shared.
- **Links to the IPS:** Local health economies will need to take into account the potential for gain/loss sharing and the IPS to overlap and either undermine or reinforce the incentives created by each scheme.

Each of these factors will inform the way shares are allocated across a local health economy. In essence, these factors will determine the rules under which the scheme will operate. Further detail on some of the possible ways gains or losses can be shared, alongside illustrative examples, are given in Section 6.5.

#### 6.4.6. Step 5: Calculating and allocating gains or losses

Once it's been decided which services and activities will form part of the arrangement, how any gains/losses will be measured and how they will be paid/recouped needs to be agreed. To carry out this calculation a number of prospectively agreed rules and principles will need to be in place, for example:

- **Metrics to use:** The basis for measuring, reporting and valuing activity will need to be agreed. For acute activity, this should utilise relevant HRGs and their respective national or locally determined prices. For other services, locally specified currencies and/or prices currently in use should be used. It's unlikely that existing or new services with no nationally or locally specified currencies and prices and in the absence of provider cost data can be included in the gain/loss sharing arrangement.
- **Monitoring and reporting:** Once metrics are agreed, local areas should decide the frequency of ongoing provider reporting and commissioner monitoring of the data, eg monthly. Infrastructure to enable the required flows of data must also be in place.
- **Payment:** The frequency of payment/recuperation of gains/losses should be determined (eg monthly), together with how payments will be made (eg through existing bilateral contracts between commissioner and providers).
- **How any gains are used:** It may be desirable to establish rules or principles for how any gains are used, for example whether they should (in full or in part) be reinvested in service transformation.

#### 6.4.7. Step 6: Any necessary adjustments

The organisations involved in the gain/loss sharing arrangement will need to agree how long it should run for and whether any of the principles agreed in Step 4 should change over this period. For example, it may be appropriate for provider shares to change over time as organisations and the new care model mature.

Over the course of the agreement, as circumstances change and new data becomes available, it may be necessary to adjust some of the agreed principles. For example, local areas will need to reassess planned activity projections when changes are made to the WPB forecast (as set out in Chapter 4). These changes may be driven by alterations to the care model or as new information becomes available. As a result, establishing a process to periodically review the agreement will be beneficial, with it also being useful to set out the criteria for when a review can trigger a change to the originally agreed principles.

### 6.5. Design parameters

This section gives guidance on how to approach designing the specific gain/loss share parameters.

#### 6.5.1. Duration of gain/loss sharing arrangement

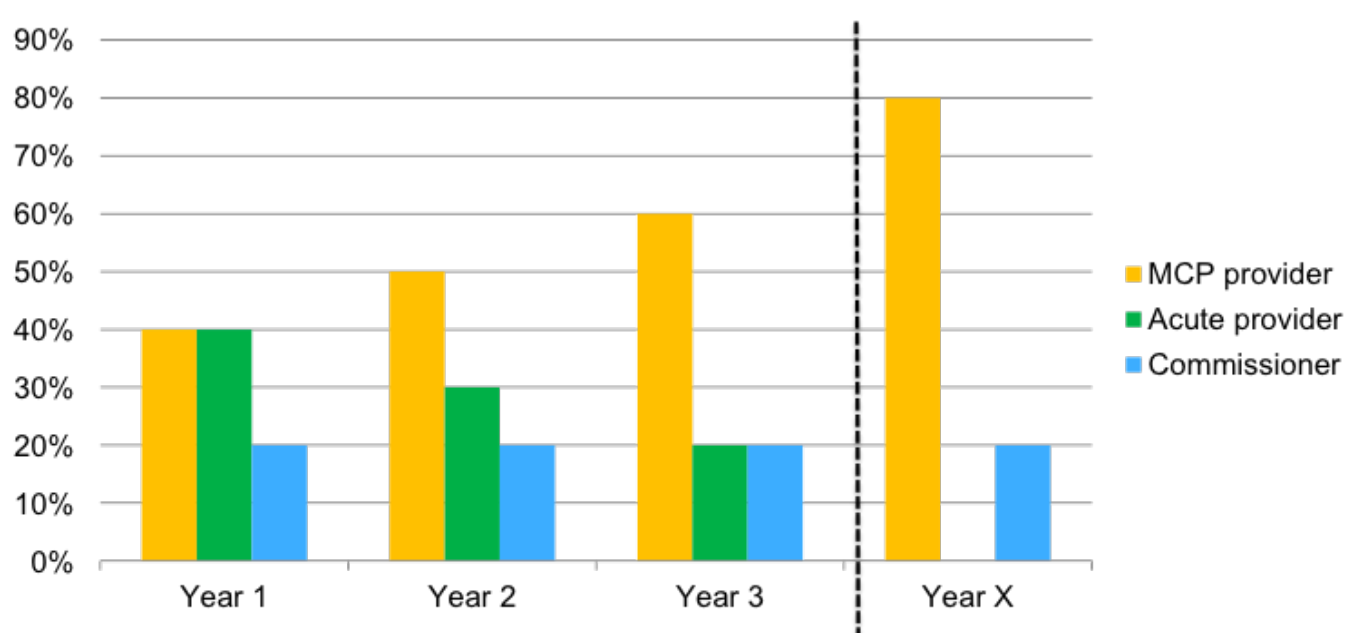
Gain/loss sharing to support the implementation of ACO care models should be multiyear. This will allow time for the mechanism to help align financial incentives and to manage the transfer of utilisation risk from commissioner to provider over a number of years. The decision on how long

a gain/loss arrangement should be guided by the overarching objectives and principles already agreed. As explained below, it may not be possible to determine the duration of the gain/loss sharing arrangement at the start of the new care model contract, with the decision on when to end the arrangement potentially being made during the life of the contract as more information and understanding is obtained.

## Aligning financial incentives

Figure 22 below illustrates gain sharing used over time to support the implementation of an MCP. The example focuses on gains only for simplicity and could be extended to the sharing of losses as well.

**Figure 22: Gain share trajectory over multiple years**



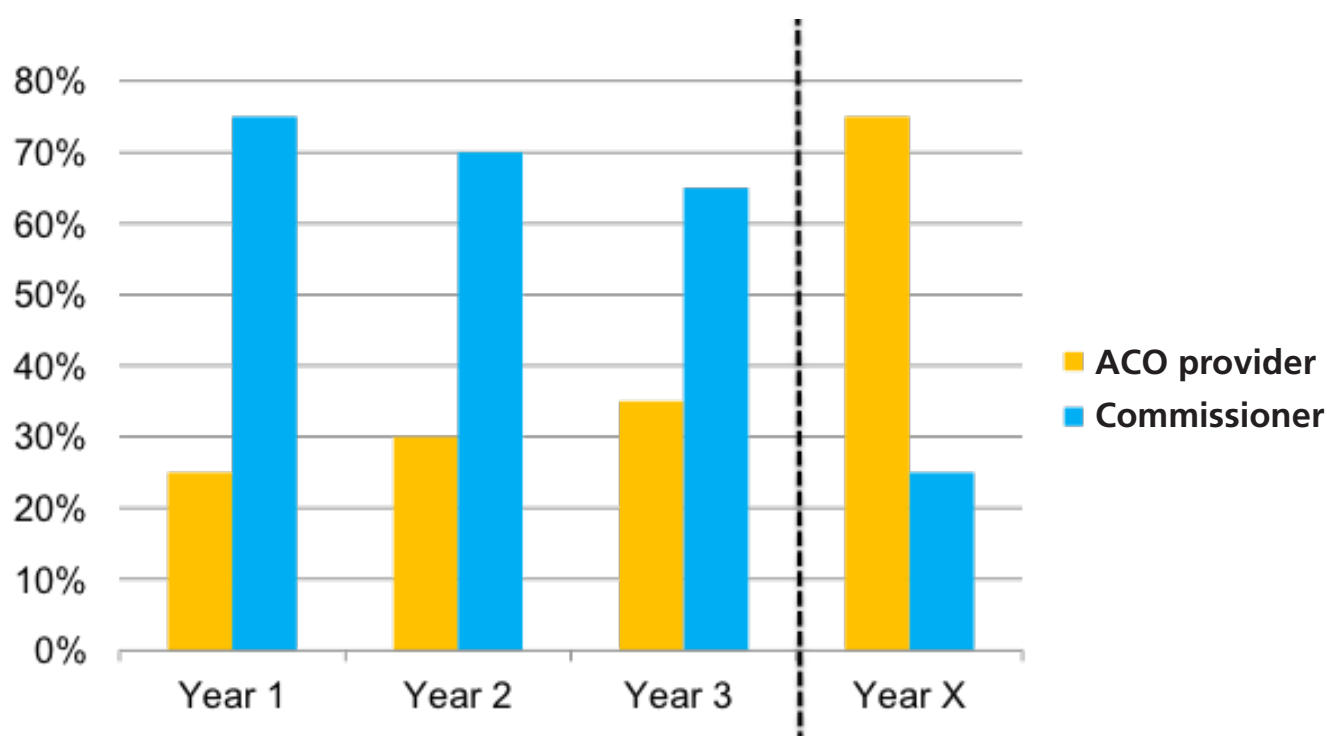
The acute provider's share of gains decreases over time to zero, reflecting their potential to reduce capacity as activity falls. In this instance, its involvement in the gain share is shorter than the overall duration of the whole contract.

The MCP provider's share increases over time, giving it an added incentive to continue to expand its services over and above the expected WPB level to drive further reductions in acute activity/spend where clinically appropriate. The duration of the gain share matches the duration of the MCP provider's contract, recognising the ongoing benefits of rewarding the MCP for its continued success in delivering improved care and system efficiencies.

## Managing the transfer and impact of utilisation risk

Figure 23 below illustrates the use of loss sharing over time to help manage the gradual transfer of utilisation risk from commissioner to provider and facilitate ACO implementation. The example focuses on losses only for simplicity and could be extended to sharing of gains as well.

**Figure 23: Loss share trajectory over multiple years**



In the initial years, the ACO provider shares a large proportion of any losses against the WPB baseline with the commissioner, recognising that the new entity will need time to establish itself. Over time, as the ACO provider scales up and builds capability, its exposure to losses can be slowly increased, and the commissioner's share reduced.

If the ACO provider is expected to mature relatively quickly, the duration of the loss share may be shorter than the duration of the contract, with risk being fully transferred to the provider by, for example, the end of year 6 of a ten-year ACO contract.

### 6.5.2. Further detail on how gains/losses will be shared

A number of factors influence the appropriate share of gains/losses between commissioner(s) and provider(s). These have been highlighted in Section 6.4.5 and are discussed in more detail below. It should be noted that how the commissioner share is distributed between a number of commissioners, and how the provider share is distributed between a number of providers within an integrated care organisation, is not covered here and is for local determination.

### Likelihood of generating gains/losses

How gains/losses are shared between commissioner(s) and provider(s) should take into account the likelihood of gains/losses being generated. This will in turn depend on how the baseline has been set, which will have considered: the extent to which envisaged new care model impacts have been incorporated into WPB forecasts, historical/current levels of volatility of demand for the service(s), and evidence from other local areas or international evidence on the potential impact of service transformation on activity shifts across providers and care settings.

If a relatively achievable baseline has been set, such that the probability of generating gains is similar to the probability of generating losses, as illustrated in Figure 24A, commissioners and



providers may decide to share similar proportions of any gains/losses made against the baseline. If a very stretching baseline has been set, so that the probability of generating losses is greater than the probability of generating gains, as illustrated in Figure 24B, local areas may wish to consider a higher share of gains than losses for the provider. Figure 24C illustrates a scenario where the agreed baseline may be deemed achievable, but uncertainty around other factors that may impact on activity and associated spend levels results in a probability skewed towards losses being generated rather than gains. Here a higher share of gains than losses for the provider may be preferred to symmetrical sharing by the provider of gains/losses.

## Ability to influence risk

Gain/loss shares should be aligned to each party's ability to influence the utilisation risk associated with specific services and/or according to how parties are impacted by those risks. This means that, as far as possible, shares should be more sophisticated than simple 50/50 shares or 'in proportion to revenue' shares. Local areas should analyse expected new care model impacts, through logic models for example, to determine which parties have the greatest levers to influence risks and which parties are likely to be most heavily impacted by those risks. The case study below shows how gains may be shared according to different parties' ability to influence risk.

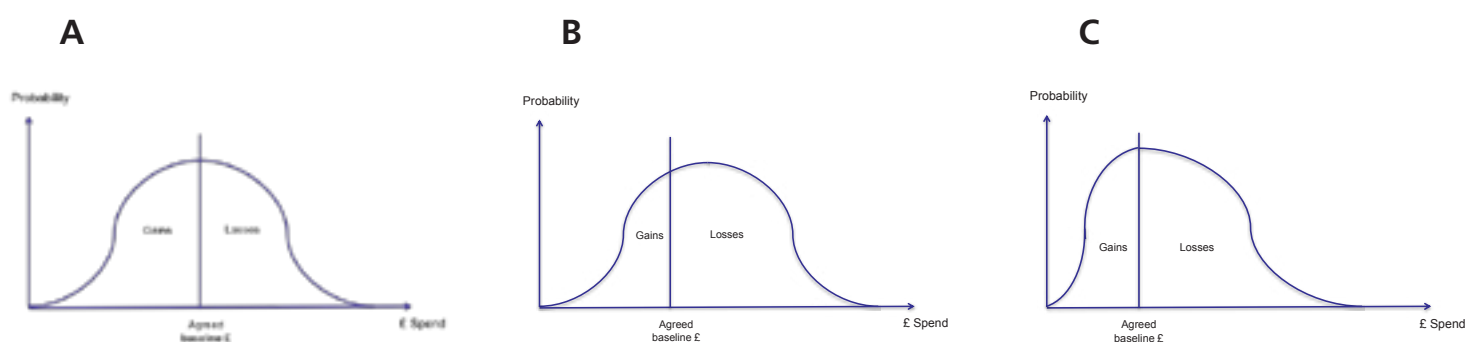
## Ability to bear risk

When determining how any risk(s) should be shared, each organisation's ability to manage risk and cope with potential losses should be taken into account. For example, the expected efficiency ask of the new care model provider in relation to historical achievement of cost efficiencies could be considered. The ISAP will be testing whether commissioners have adequately assessed as part of the procurement the ability of the preferred provider to take on the risks associated with the proposed contract.

## Appetite to bear risk

Appropriate risk shares between parties will also depend on their willingness to bear risk. For example, a provider that is establishing a new care model may be risk averse and have little appetite to bear downside risk, particularly at the start of the contract. However, as the provider scales up, it may be willing to accept a higher percentage of any losses if it's able to keep a greater amount of any gains.

**Figure 24: Probability scenarios for generating gains and losses. (A) Probability of generating gains and losses is equal; (B) probability of generating losses is greater than the probability of generating gains; (C) skewed probability creating greater likelihood of losses being generated rather than gains**



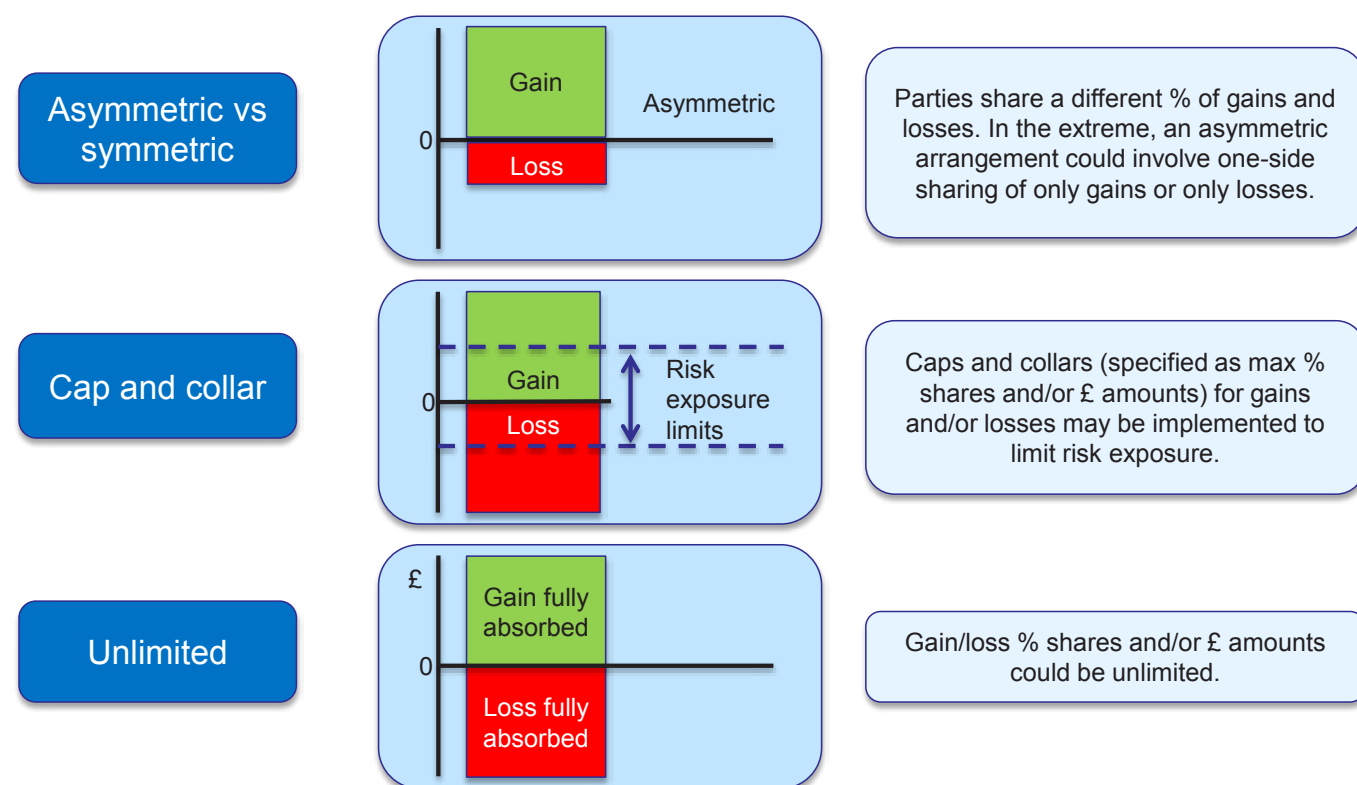


## Risk exposure over time

The level of risk exposure (both downside (losses) and upside (gains)), will likely need to change over the duration of the gain/loss sharing arrangement as the provider(s) of the new care model becomes established and organisations mature. Over time, these organisations should improve their ability to manage utilisation risk and this should be reflected in the arrangement.

Figure 25 illustrates some of the gain/loss design parameters that may need to be varied over time to manage levels of risk exposure.

**Figure 25: Gain/loss parameters that may need to be varied over time**



Given the current financial climate and historical outturn performance against expected/planned performance (on efficiency, outcome/performance metrics, etc), local areas should consider implementing one-sided gain sharing in the initial year(s) of the new care model contract. This should ideally transition to two-sided gain/loss sharing, with potential asymmetric sharing (higher share of gains than losses) in the near term, recognising the high levels of uncertainty around new care model impacts. This is particularly important if a very stretching baseline has been set, such that the probability of generating losses is greater than the probability of generating gains.

In the later years of the new care model contract, the provider may take on significantly higher percentage shares of both gains and losses once it has established new patterns of care delivery and adjusted to new patient flows.

Caps and collars should be used in the transition to limit excessive exposure to both upside and downside risks. Over time, once providers have built capacity to manage and absorb risk, these limits can be removed so that share of gains/losses is not limited.

## Improvement Payment Scheme

In deciding whether and to what extent gain/loss shares should be linked to IPS indicators, local economies will need to take into account the impact this will have on a provider's ability to share in any gains/losses. For example, if linking a provider's share of gains to IPS indicators means that the provider is likely to share a very minimal level of gains, the intended incentive signals from the gain/loss mechanism will be dampened.

If possible, measures used for gain/loss arrangements and those used for IPS should be distinct but aligned so that the incentives created by each are balanced and do not conflict. This will mitigate the risk of duplication and possible weakening of the financial incentive signals from the two mechanisms.

## Minimum savings rate

It may be appropriate to establish a minimum savings rate, with only gains/losses over and above a minimum level being shared. This recognises that estimating baselines accurately is difficult and any gains/losses generated relative to a baseline may be due to random variation.

### Key lessons and principles

Based on evidence and learning from international schemes (eg shared savings schemes for accountable care organisations in the USA), it seems appropriate for multiyear gain/loss sharing arrangements to be one-sided to begin with; for example, gains shared only in the initial year(s) of the new care model contract. Equally, there should be a clear plan to move to two-sided arrangements, ie gain/loss sharing as soon as is feasible.

Proportionality should guide the determination of percentage gain/loss shares, with shares informed by the likelihood of gains/losses being generated against baseline, as well as aligned to parties' ability to influence specific risks and/or according to how they are impacted by those risks.

## 6.6. Next steps

Initially, as ACO providers will have limited capacity to manage and absorb risk, it will be important to shadow test gain/loss sharing arrangements to better understand their potential impacts and to ensure that design of the gain/loss parameters will deliver the intended objectives.

The next publication of this handbook will provide further detail to help commissioners and providers to develop gain/loss sharing arrangements locally. This will be informed by the ongoing work by MCP and PACS as they continue to develop WPBs and gain/loss sharing, working towards shadow testing and implementation within contracts.

The NHS Five Year Forward View sets out a vision for the future of the NHS. It was developed by the partner organisations that deliver and oversee health and care services including:

- NHS England
- Care Quality Commission
- Health Education England
- The National Institute for Health and Care Excellence
- NHS Improvement
- Public Health England

**Our values:** clinical engagement, patient involvement, local ownership, national support