Ref: care.data/Programme Board/Paper 04
Title: **Information Analysis Linked Primary Care – Secondary Care Dataset**
Author: Trevor Anders
Programme Board Sponsor: Eve Roodhouse, Programme Director
Purpose: To provide the Programme Board with an update on the current position and plans for progression of analysis of the linked data available during the pathfinder stage
Background: Plans for the analysis of the linked data set were originally drawn up in late 2013 prior to the decision to delay the start of the extraction of primary care data from general practices. Since then changes have been made to the original analysis proposals to take account of developments since the announcement of the listening exercise.
Key Points: Analysis by NHS England to support evidencing of the GPES Benefits Plan will continue as originally planned. A new primary care function within the HSCIC will also work alongside NHS England colleagues and will develop its own complementary analysis plan which is expected mainly to be focused around aspects of data quality.
Desired outcome(s): For discussion
Circulation: Programme Board attendees and Advisory Group

Purpose

1. This note is intended to inform the Programme Board of the:
    i.   Current proposals regards analysis of the linked primary-secondary care data;
    ii. Plans for progression of the analysis work; and
    iii. To seek feedback and support for the plans.

Background

2. As part of the process of agreeing the extraction of primary care data through the General Practice Extraction Service (GPES) a Customer Requirement Summary (CRS) and Benefits Plan were developed. These are key documents which are part of the GPES governance model and support the submission of extract requests through the GPES Independent Advisory Group (IAG). Copies of the documentation for care.data can be found at http://www.hscic.gov.uk/gpes/caredata.

3. At the time these documents were produced in early 2013 it was agreed that access to the linked data set should be restricted to requests to support commissioning purposes and therefore the Benefits Plan was constructed to evidence the kind of questions and analysis that it was envisaged may be possible with the linked data set relating to commissioning purposes.

4. At the time this was being progressed the HSCIC did not have an in-house analysis function with the expertise to analyse primary care data and it was agreed that early access should be facilitated for the NHS England analytics team to access the linked data. It was also agreed that this analysis should be focused on evidencing the benefits as set out in the IAG Benefits Plan to address IAG concerns regards proving the utility of the data within a clearly defined remit before expanding access to the data for wider purposes.

5. To support this the programme worked with the NHS England analytics team and agreed a scope for the initial analysis of the data, the scope of this is detailed in Appendix 1. It was planned that this initial analysis would be for a limited period after which access to the linked data set would have been made available through the standard HSCIC data extract and linkage applications process.

Plans

6. Following the announcement of the pause in the roll out of care.data for primary care and the subsequent listening period there have been additional developments which have changed the original plans.

7. The HSCIC has been developing its own in-house Primary Care function and this is expected to be operational by the end of September. In addition the programme has agreed a set of further controls for the pathfinder stage which include restricting access to the linked data set via a Secure Data Facility (SDF) within the HSCIC and that access will

only be available to HSCIC, NHS England, Care Quality Commission (CQC) and Public Health England (PHE) along with the pathfinder sites[1].

8. Renewed discussions with the NHS England analytics team have confirmed that they will continue with the previously agreed scope of analysis work to support the IAG benefits plan. It is intended that the HSCIC will look at analysis of the data mainly from a data quality basis and currently the team are determining the legal position regards the HSCIC role and whether a set of experimental statistics are required to be developed.

9. The Data Workstream within the programme is planning a workshop for late October to bring together the HSCIC and NHS England teams to explore further how they might work together and the respective scope each should take in terms of types of analysis of the data.

10. It is also intended that once the pathfinders have been formally agreed that the HSCIC will work with the pathfinder CCGs and practices to understand their information reporting needs to determine what outputs could/should possibly be derived and produced from the centre to support the pathfinders and to better understand the local information reporting needs.

11. An initial meeting has also taken place with CQC to determine what their needs are and to quality assure the requirements for the SDF. Further meetings will be scheduled through the HSCIC account relationship managers with both CQC and PHE and they will be invited to a further workshop with both NHS England analytics and HSCIC primary care team. The objectives for the workshop will be to agree the scope and boundaries that each organisation will work to and how they can most usefully collaborate on the analysis to ensure shared learning and understanding of the data and any potential limitations they should take into account as part of drawing conclusions/findings from the data.

Risks and Issues

12. Early work in reviewing the agreed data extraction with clinical leads, NHS England analytics, Clinical Commissioning Groups (CCGs) and Commissioning Support Units (CSUs) has identified there will be a number of limitations in the data such as the lack of any historic data, the inability to easily identify Long Term Conditions, the volume of data and the general breadth of the data extracted which will limit the utility of the data and the analysis that can be conducted particularly in the first few months when there will be very limited history. Therefore any conclusions or early findings resulting from the analysis will need to be treated with caution and are unlikely to be statistically valid.

13. The limitations in the data are likely to impact on what can be evidenced to support the statements made in the IAG benefits plan though it is recognised that this lack of

---

[1] CQC & PHE access and usage will be determined by their statutory role and the obligations on the HSCIC in providing data to meet those statutory requirements under the Health & Social Care Act. Usage of the data beyond meeting those needs specifically set in legislation will be subject to the recommendations made by IAG in response to the submission made to them on 11th September re allowing access to the data for purposes beyond commissioning.

evidence may reinforce and support any later requests to make changes in the primary care extract specification as part of the proposed roadmap development following the pathfinder stage.

14. Although it has been agreed that access to the data for the pathfinder stage should be through the SDF there is a likelihood that pathfinder CCGs, their practices and the CSUs that support them may want/need more direct access to the linked data outside of the SDF. These requirements will be better understood once there has been the opportunity for dialogue with the selected pathfinder sites.

15. It is possible that the HSCIC primary care function will not be able to fulfil all the reporting needs that pathfinders may request and therefore alternative ways to address these needs such as through the CSUs may need to be explored.

Actions Required

16. This paper is mainly provided for information however the Board are invited to comment on the outline plans and risks identified.

Appendix 1 – Care.data NHS England Analysis Questions

No.	IAG Benefit	Benefit Scenarios	Question	Example / Description in Plain English	Technical Description	CDS Type
1	Monitoring Outcomes	Determining effectiveness of treatment	Analyse the patients that were discharged in HES in Month 1 & 2, and subsequently the primary care data to look at what interactions these patients had with their GP / Practice. In addition, if these patients were readmitted to hospital (readmissions in HES) By Practice / CCG, Gender and Age.	What contact do people have with their Practice after having been in hospital, and how does this affect their chance of being readmitted.	Select date of discharge within range and select where GP appointment date was greater than this range. Select HES data by primary diagnosis/procedure and select diagnostic information from GPES too. Can't do readmissions using national algorithm but could set up using our own simplified algorithm?	APC
		Evaluation of re-ablement and discharge	Due to the limitations of Phase1 of the care.data extract, this benefit is not achievable in this phase.	N/A		
		Monitoring a care pathway or Finding the cost of a treatment or pathway	Analyse the Outpatient Appointments in HES, Month 1 & 2 where patients had a follow up appointment and subsequently the primary care data to look at what interactions these patients had with their GP / Practice. By Practice / CCG / Location, Gender, Age and Diagnosis Provider	What contacts do people have with the Practice when they have had a recent outpatient contact.	As above but using outpatients data instead (Selecting only attendances - not all appointments)	OP

2	Improved Performance Monitoring	Identification of A&E patients who are high risk	Patients who have had an A&E attendance and non–elective admission in HES – calculate in primary care data how many times these patients saw their GP Practice or had a Practice Contact beforehand in month 1 & 2. In addition, analyse the following: Where the patient has had a Primary Care Contact, calculate: a. from primary care dataset - the contact type such as GP, Nurse, Repeat Prescriptions etc.; and b. from HES dataset – Primary Diagnosis, Provider and Length of stay By Practice / CCG, Gender and Age.	What sort of contact did people have with their Practice before they had an emergency event with secondary care. Could use to see if certain types of admissions could have been avoidable with primary care intervention beforehand.	Select date of GP contact within a range and select attendance or emergency admission date is greater than GP contact date. Split by diagnosis, procedure, etc. Probably easer to merge this and the one below to look at all admissions by admimeth	AE & APC
			As above but for Elective Admissions	As above	As above but for Elective Admissions	APC
		Monitoring referral rates	In Primary Care Data, analyse the referral details and characteristic of patients such as Age, Diagnosis etc by Practice / CCG Level and compare this analysis against the QOF Register to identify the patients that were referred to hospital for an Elective Admission. In the primary care data, analyse the referral rates by Practice / CCG Level.	Does the data from primary care and HES coincide and are they consistent?	If referred from GP to hospital were they then admitted to hospital (elective). Select date of GP contact within a range and select elective admission date is greater than GP contact date. Compare to QOF seperately	APC
		Monitoring referral rates	Which types of referrals do not result in an appointment and why? Is there consistency in the rate (perhaps implying an inevitable drop-out) or big inconsistencies that might imply poor			

			patient experience?		
		Monitoring and benchmarking admissions	For patients with records both in Primary Care and HES dataset who are admitted to hospital, sort the data by Ethnicity for patients of certain conditions (diagnosis).	Is ethnicity consistently recorded between primary and secondary care?	Find all patients with a record in HES (do they mean APC/OP/AE or just APC?) and a record in GPES and output where ethnicity codes do not match	TBC
		Monitoring and benchmarking admissions	Where are the biggest health inequalities and how does access to primary care services vary?	How are we doing on tackling health inequalities? For example, what is the level of seasonal flu vaccination uptake for hard to reach populations, e.g. people in care homes, etc.		
3	Improved management of patients with QOF conditions	Find numbers of patients with a specific disease	Calculate the number of patients by GP Practice, Condition (Diagnosis), Characteristic (Gender, Age) and the associated prescribing information.	What are the prescribing rates for different conditions and how do they vary between practices?		
		Evaluation of needs analysis monitoring	In the primary care data, look at the prescribing information or health feature in QOF such as Diabetes, and analyse the characteristic of the patients that does not meet the QOF criteria.	Are there drugs prescribed for patients that would indicate a condition but that patient is not on the QOF register? E.g. insulin prescription but not on diabetes register.		
		Evaluation of needs analysis monitoring	Can we further disaggregate data on QOF exceptions, in particular discretionary vs. non-discretionary exceptions?	Are QOF exceptions being applied appropriately?		

4	Earlier Diagnosis of patients	Identify numbers of patients receiving prescriptions but who do not subsequently collect them	In the primary care data, investigate the quality of data to analyse if the vaccinations / immunisations and screening events are recorded consistently. By GP Practice / CCG	Are vaccinations / immunisations and screening events recorded consistently?		
			Analyse patients who were regularly attending primary care before reaching secondary care (inc. via A&E) and being diagnosed with serious conditions (cancer, etc.).	Missed diagnoses in primary care		
			In the primary care data, analyse the patients on disease registers with HES diagnosis codes and identify potentially 'missed diagnoses' in primary care. For example Dementia as a starting point given that the definition is relatively well established (in terms of which codes).	Missed diagnoses in primary care		
5	Improving commissioning of primary care activity	Performance monitoring	Analyse both the primary care and HES dataset, to look at the variation in acitvity across the primary and secondary care for number of events recorded per patient to analyse the distribution. By GP Practice, CCG, Age, Gender and Diagnosis	N/A	Find all patients with a record in HES (do they mean APC/OP/AE or just APC?) and a record in GPES and provide a count by NHS number so ratios could be calculated	
		Performance monitoring	Analyse how GP workload caries according to the characteristics of patients. Understand how the demographic characteristics and morbidity of patients impacts on use of secondary care	What drives GP workload? How can the allocations formulae be improved to better reflect the drivers of health care need?		
		Performance monitoring	What non-OP services are GPs referring into? For which groups of patients?			

		Performance monitoring	How frequently are GPs referring for diagnostic tests and how many of these are captured in HES?			
		Performance monitoring	Analyse condition-specific pathways and co-morbidities, in particular for patients being seen in primary care but not being referred to secondary care.	How well are patients being managed at home and inappropriate referrals being avoided? How does this vary by practice, disease, etc.		
		Monitoring new community initiatives	Due to the limitations of Phase1 of the care.data extract, this benefit is not achievable in this phase.	N/A		
6	Improved Data Quality	N/A	Analyse the data completeness in primary care dataset by following: a. Age; b. Ethnicity; c. Gender; and d. Small Area of Residence By Practice / CCG, National and Regional Level	N/A	This sounds like analysis of the GPES data on it's own (is this 'allowed' under the rules?)	
			Analyse the patients that were discharged in HES in Month 1 & 2, and subsequently the primary care data from Month 1 to 4, to look at if these patients have had their diagnosis data added to their record.	If patient have been in hospital with a particular diagnosis, it would be expected that this would be noted in the primary care data when they next saw their GP	Find all patients with a record in HES (do they mean APC/OP/AE or just APC?) and pull out diagnostic data for both to check a) it is present and, b) that is corresponds. This might involve record level analysis which would be time consuming or we might have to map READ codes to ICD10 codes...	TBC