Digital Health and Care Wales

Data Quality Status Report
APC Data Set

Document Version	1.0

Status	Approved

Document author:	Toni Price
Approved by	Daniel Hughes
Date approved:	11/08/2021
Review date:	

TŶ GLAN-YR-AFON 21 Heol Ddwyreiniol Y Bont-Faen, Caerdydd CF11 9AD

TŶ GLAN-YR-AFON 21 Cowbridge Road East, Cardiff CF11 9AD

TABLE OF CONTENTS

1 Data Set

1.1 TITLE

Admitted Patient Care data set (APC ds)

1.2 SPONSOR

Welsh Government (WG)

1.3 IMPLEMENTATION DATE

1st April 1999

1.4 CHANGE HISTORY

See NHS Wales Data Dictionary[1]

1.5 DATA SET PURPOSE

The APC data set is the principal source of secondary use data for hospital admissions. Each record in the data set describes a Finished Consultant Episode (FCE). The data set has a wide range of uses including:

- Service improvement
- Hospital capacity planning
- Budget planning
- Financial costing
- Performance monitoring
- Public health surveillance

This data is submitted by each provider organisation to the NHS Wales Informatics Service's Information Services Division (ISD), where the data is loaded into the national database. At the end of each year, the data for that year is 'frozen' to ensure that National Statistics for that year remains unaffected by changes to the database as a result of data resubmissions.

[1] http://www.datadictionary.wales.nhs.uk/WordDocuments/outpatientdatasetopds.htm

2 Document Purpose

2.1 FUNCTION

ISD provides a central data processing, analysis and publishing service for NHS Wales. A key element of this process is to ensure that the data being processed is of suitable quality to maintain the integrity of the database which, in turn, enables the reporting of meaningful health information.

This document describes a range of data quality issues affecting this data set.

2.2 AUDIENCE

This document acts as a situation report for the Sponsor as well as an information resource for other stakeholders who base decisions on the accuracy of this data.

2.3 FREQUENCY

The document is issued annually to accompany annual publications and data releases.

2.4 INFORMATION SOURCE

The aim is to describe the quality of the data held centrally in the Digital Health and Care Wales national database. The Data Quality and Data Acquisitions teams within ISD are in regular contact with the health boards who supply this data, as well as the WPAS[2] (Welsh Patient Administration System) development team to ensure that the data being entered on hospital systems and extracted from them comply with the standards set out in the NHS Wales Data Dictionary and relevant Data Set Change Notices (DSCNs)[3].

2.5 SCOPE

The following set of data quality dimensions are covered in this report:

- **Data Validity**. The term 'data validity' refers to whether the submitted data has been provided in the agreed format and, where applicable, whether it is populated with a nationally-agreed value.
- **Data Consistency** refers to whether related data items within the same data set are consistent with one another. For example, a record that indicates a female patient has undergone a prostate biopsy should be considered inconsistent and would require investigation.
- **Data Timeliness** is simply a measure of whether the data file was submitted in accordance with national timescales.
- **Data Completeness**. In this case, this is a measure of the ratio of records submitted : records loaded.

These are fundamental to the quality of the data which is submitted and, in turn, processed through to the national database.

Aside from the data quality dimensions listed above, this document does not seek to review the accuracy of the data reported via the APC ds – i.e. whether reported activity is a true reflection of the activity being carried out within NHS Wales organisations.

[2] WPAS was formerly called Myrddin

[3] All new DSCNs are published on the Digital Health and Care Wales Data Standards website via: Data Standards - Digital Health and Care Wales (nhs.wales)

Note also that nationally defined default or bucket codes are permitted and are therefore classed as valid values.

Further information about these dimensions can be found on the Digital Health and Care Wales Data Quality website[4].

3 Data Quality Standards

3.1 VALIDATION AT SOURCE SERVICE (VASS) CHECKS

This data set is used for high profile National Statistics where a high level of quality assurance is required. VASS provides an online resource for submitting organisations to check the quality of their data before formally submitting it to ISD to be processed through to the national database. VASS is comprised of 3 main types of data quality checks as described below:

- **Data Load checks** are used to protect the integrity of the database by identifying invalid values within a record. If a data load error is triggered, the whole record is rejected by the system, preventing it from being processed through to the national database. The fact that load errors prevent records from being loaded means that these are often reviewed and resubmitted immediately. While this has been a successful method of maintaining the quality of this data set, it is reliant on the cooperation of the data provider in reviewing these errors promptly.
- **A Data Validity check** tests whether the recorded entry within the associated database field is a valid national value. These national values are defined in the NHS Wales Data Dictionary and lists of codes are available from the Welsh Reference Data Service[5]. Data Validity checks have been in operation since April 2010.
- Some data items are interdependent. For example, a patient's date of birth must not be after their attendance date. Relationships between data items are checked using **Data Consistency checks**. These were introduced for APC in April 2009.

These checks are reviewed and updated as necessary.

3.2 REGULAR MONITORING

Data Validity and Consistency performance is monitored on a monthly basis. The **Data Quality Standards** that each data provider must adhere to are defined by sets of indicators and nationally-agreed targets. These are based on the aforementioned VASS checks. Data Validity and Consistency reports are used to measure compliance with these standards

Further information on Data Quality Standards and how the quality of data is monitored can be found on the Digital Health and Care Wales Data Quality internet site[6]. The reports themselves are published on the corresponding intranet site[7].

[4] Data Quality - Digital Health and Care Wales (nhs.wales)
[5] http://wrds.wales.nhs.uk (accessible to NHS Wales users only)
[6] Data Standards - Digital Health and Care Wales (nhs.wales)
[7] http://nww.nwisinformationstandards.wales.nhs.uk/data-quality (accessible to NHS Wales users only)

4 Data Set Quality Status

4.1 DATA VALIDITY

Regular monitoring and provider cooperation mean that data validity is generally high. A copy of the annual Data Validity report for 2020-21 is shown in Appendix A. Issues causing percentages to fall considerably below the target (>4%) are explained and resolved by the health boards where resource and system constraints permit.

The validity targets for **Main Specialty (Consultant)** and **Speciality of Treatment Code** are being met every year by all organisations apart from Powys. The records that are causing this low validity mainly relate to activity where a general practitioner is responsible for the patient during their inpatient stay and/or where the patient is treated under the specialty of general practice. The main specialty code relating to GPs changed when the list of values for this data item was revised in April 2015 as per DSCN 2014 / 07[8] and the specialty of treatment codes for general practice ceased to be valid in April 2016 as per DSCN 2014/08[9].

Three other organisations also submit activity where GPs are responsible for patients during their inpatient stays and/or where patients are treated under a GP specialty. They also use the codes that are no longer valid. The validity targets are still met in those three organisations, as although there are a similar number of invalid codes, the proportion is negligible due to their larger total volume of records. This issue has been highlighted in Powys as it accounts for a high proportion of their overall activity. An upgrade to the WPAS system in Powys is needed before the validity of these data items will improve.

4.2 DATA CONSISTENCY

A full breakdown of Data Consistency performance is shown in Appendix B. Data Consistency compliance is good in general.

The indicators that are showing low values are all indicators where the denominator in the calculation is a subset of the total number of records. As the calculations are based on a relatively small number of records, the percentages can be somewhat deceptive. In BCU for example, the low consistency of **Discharge Method v Specialty (of Treatment)** relates to only 4 records, while that for **Discharge Method vs. Discharge Date & Date of Birth (i.e. Age)** relates to 4 records across all Health Boards. Velindre has a low consistency rate for **Decision to Admit Date vs. Waiting List Date** as they do not record an inpatient waiting list, both the 'Waiting List Date' and the 'Decision to Admit Date' are calculated rather then entered into Canisc.

4.3 DATA TIMELINESS

Issues with timeliness are rare due to an established process of file submission and sign-off via the NHS Wales Data Switching Service (NWDSS). The Data Acquisitions team issue reminders to data providers ahead of the monthly submission deadline and provide assistance with any VASS errors to reduce delays and minimise the probability of missed deadlines. There was only one late submission for APC data in 2020-21 which was due to technical issues.

In addition to monthly deadlines, there is an annual deadline for resubmissions (15th June 2021) which

[8] http://www.nwisinformationstandards.wales.nhs.uk/opendoc/253025
[9] http://www.nwisinformationstandards.wales.nhs.uk/opendoc/258253

allows providers to improve the quality of their data before it is frozen. All health boards met the deadline.

4.4 DATA COMPLETENESS

ISD data processing timescales must be adhered to in order to ensure compliance with reporting deadlines. If a monthly submission deadline is missed, the data cannot be processed until after the submission deadline for the following month. This can result in temporary data completeness issues. This does not affect the data used in annual reports as these are only run after files for the entire year have been received (and resubmitted where necessary) using the frozen data.

With the existence of Data Load checks there is an added risk of data completeness issues if invalid data is submitted. Although rejected records are generally reviewed and resubmitted before the data is loaded, if these are not corrected, the national database (and any reporting outputs) will contain incomplete data. This is not a significant issue at present as a relatively small number of records are rejected by the system each month and not loaded into the national database. Any instances where a high proportion of records are rejected are flagged up on Data Completeness reports. These are monitored by the Data Acquisitions team and issues are communicated to the submitting organisation immediately requesting that the data is resubmitted in time for the data to be processed.

Three organisations (BCUHB, Powys and Swansea) met both clinical coding completeness targets while another three organisations (Aneurin Bevan, Cwm Taf and Hywel Dda) failed to meet either of them. Aneurin Bevan (92.8%), Cwm Taf (89.1%) and Hywel Dda (85.9%) failed to achieve the national standard of 95% coded within 3 months of episode end date, while Aneurin Bevan (92.8%), Cardiff & Vale (96.8%), Cwm Taf (89.1%), Hywel Dda (85.9%) and Velindre (97.6%) failed to achieve the 98% target for rolling 12 months' data.

4.5 ADDITIONAL ISSUES

The data quality dimensions described above capture the major issues which can be easily monitored. However, there are some additional issues which, although not captured by regular monitoring, are highlighted to the Data Quality team on an ad hoc basis. The table in Appendix C describes the current position.

For further information regarding these issues, please contact the Data Quality team via data.quality@wales.nhs.uk.

4.6 QUALITY ASSURANCE

ISD follows a routine process to assure the quality of the data used in National Statistics. This process is described in the document *Data Quality Assurance – National Statistics (June 2014*, which is available from the Digital Health and Care Wales Data Quality Team on request*)*.

4.7 IMPACT ON REPORTING AND PUBLISHING

There are no major issues preventing this data from being used for reporting, providing that the recipient is made aware of the relevant issues described in this report.

4.8 OVERALL DATA QUALITY STATUS

Well established processes for submitting, checking and monitoring the quality of this data set means that the timeliness, completeness, validity and consistency of the data are generally good. These dimensions

continue to be monitored on a regular basis to further improve quality.

Improvements to the other areas summarised in Appendix C are largely dependent on developments to operational systems or to the structure and scope of the data set itself.

WHC (2015) 027[10] was issued in June 2015 to introduce a national initiative to address the causes of poor information quality. The new initiative, namely the Information Quality Improvement (IQI) initiative, has now been established and a work programme has also been developed along with detailed proposals for tackling the underlying causes of a number of issues affecting information quality. However, due to COVID-19 this initiative was put on hold as members of the group had working tasks which took priority over this. Once workloads have resumed, this group will then look to meet on a regular basis again.

More information about the initiative can be found by visiting www.iqi.wales.nhs.uk.

[10] http://howis.wales.nhs.uk/doclib/WHC2015027-e.pdf

5 Appendix A: APC Data Validity Report 2020-21

Data Item	DATA VALIDITY STANDARD	All Welsh Providers	Aneurin Bevan University LHB	Betsi Cadwaladr University LHB	Cardiff & Vale University LHB	Cwm Taf Morgannwg ULHB	Hywel Dda University LHB	Powys Teaching LHB	Swansea Bay ULHB	Velindre NHS Trust
APC submission received by the 17th	-	-	✓	✓	✓	✓	✓	✓	✓	✓
Number of Records Loaded	-	801258	174401	162128	111266	93735	87170	2588	107611	62359
Administrative Category	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Admission Date	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Admission Method	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Consultant Code	98%	97.1%	94.2%	96.4%	95.5%	✓	✓	77.5%	✓	✓
Date of Birth	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Decision to Admit Date	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Discharge Date	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Discharge Destination	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Discharge Method	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Episode Start Date	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic Group	98%	✓	97.8%	✓	✓	✓	✓	✓	✓	✓
HRG Code†††	95%									
Intended Management	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Last Episode in Spell Indicator	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Legal Status	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Local Health Board of Residence	95%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Main Specialty (consultant)	98%	✓	✓	✓	✓	✓	✓	65.2%	✓	✓

NHS Number	95%	✓	✓	✓	✓	✓	✓	✓	✓	✓
NHS Number Status Indicator	95%	✓	✓	✓	✓	✓	✓	✓	✓	✓
NHS Number Valid & Traced	95%	✓	✓	93.2%	✓	✓	✓	✓	✓	✓
Patient Classification	95%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Postcode††††	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Principal diagnosis†	95%	✓	✓	✓	✓	✓	91.7%	✓	✓	✓
Principal procedure Code†/††	95%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Principal Procedure Date	95%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Referrer Code	98%	73.7%	36.2%	82.8%	✓	75.2%	✓	97.8%	58.0%	✓
Registered GP Practice Code	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sex	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Site Code (of Treatment)	98%	✓	✓	✓	97.7%	✓	✓	✓	✓	✓
Source of Admission	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Specialty of Treatment Code	98%	✓	✓	✓	✓	✓	✓	64.1%	✓	✓
Waiting List Date	98%	97.1%	✓	85.7%	✓	✓	✓	✓	✓	✓

† "Principal ICD Diagnosis", "Principal Procedure Code" and "HRG Code" will show as blank for the first 3 months of a new financial year.

†† "Principal Procedure Code" only reports the % validity of all records where a primary procedure code is present on an episode.

††† "HRG Code" (HRG v4) is not presently monitored for data validity as it is currently a derived field.

†††† "Postcode" may occasionally be incorrectly flagged as invalid due to issues with the postcode file received from ONS.

6 Appendix B: OP Data Consistency Report 2020-21

Data Consistency Check	DATA CONSISTENCY STANDARD	All Welsh Providers	Aneurin Bevan University LHB	Betsi Cadwaladr University LHB	Cardiff & Vale University LHB	Cwm Taf Morgannwg ULHB	Hywel Dda University LHB	Powys Teaching LHB	Swansea Bay ULHB	Velindre NHS Trust
Admission Date vs. Date of Birth	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Admission Method vs. Intended Management	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Admission Method vs. Patient Classification	95%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Admission Method vs. Source of Admission*	98%	✓	✓	97.4%	95.3%	✓	✓	✓	✓	81.0%
Decision to Admit Date vs. Admission Date	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Decision to Admit Date vs. Waiting List Date	98%	75.8%	✓	✓	✓	✓	✓	✓	✓	0.1%
Discharge Method vs. Discharge Date & Date of Birth [i.e. Age]*	98%	0.0%	n/a	0.0%	n/a	0.0%	n/a	n/a	n/a	n/a
Discharge Method vs. Discharge Destination*	98%	✓	✓	✓	✓	✓	✓	✓	✓	92.0%
Discharge Method vs. Specialty (of Treatment)*	98%	47.8%	n/a	0.0%	33.3%	33.3%	n/a	n/a	✓	n/a
Episode End Date vs. Admission Date	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Episode End Date vs. Discharge Date	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Episode End Date vs. Date of Birth	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Episode End Date vs. Episode Start Date	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Episode Start Date vs. Admission Date	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Episode Start Date vs. Discharge Date	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Episode Start Date vs. Date of Birth	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
HRG Code vs. Sex†*	95%	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Last Episode in Spell vs. Episode End Date & Discharge Date*	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Legal Status vs. Specialty (of Treatment)***	98%	91.1%	✓	✓	✓	✓	2.4%	✓	✓	n/a
Patient Classification vs. Discharge Date & Admission Date [i.e. Length of Stay]*	95%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Postcode vs. Local Health Board of Residence**	95%	✓	✓	✓	✓	✓	✓	✓	✓	91.6%
Primary Diagnosis Code vs. Admission Date & Birth Date [i.e. Age]†*	95%	✓	80.0%	93.0%	✓	✓	✓	n/a	n/a	n/a
Primary Diagnosis Code vs. Sex†*	95%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Primary Procedure Code vs. Sex†*	95%	✓	✓	✓	✓	✓	✓	✓	✓	✓

Primary Procedure Date vs. Episode Start Date & Episode End Date	95%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Referrer Code vs. Referring Organisation Code	98%	✓	✓	✓	✓	92.2%	✓	✓	✓	✓
Specialty (of Treatment) vs. Sex*	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Waiting List Date vs. Admission Date	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Waiting List Date vs. Admission Method	98%	✓	✓	✓	✓	✓	✓	✓	✓	✓

The term "data consistency" refers to whether related data items within the same data set are inconsistent. For example, a record that indicates a male patient has given birth should be considered inconsistent and would require investigation).

† The national standard for clinical coding stipulates that primary diagnosis/procedure should be completed within 3 months of episode end date. However, these checks look at all submitted coding irrespective of the elapsed time since the episode end date.

* Reported performance shows the percentage of consistent records for those records that contain a dependant value only. Please refer to the supporting documentation for a full breakdown of the dependencies for each check.

** "Postcode" may occasionally be incorrectly flagged as invalid due to issues with the postcode file received from ONS.

*** This measure shows consistency of mental health related specialties only.

n/a = no relevant activity data submitted to test in relation to this check

7 Appendix C: Additional Data Quality Issues

Issue	Impact	Proposed Resolution	Benefit	Status
Maternity 'tail' / Maternity stats	Inconsistency between APC figures and data held in local electronic maternity systems (and in the Maternity Indicators data set).	As Maternity stats are now sourced from the Maternity Indicators data set instead of from the maternity 'tail', consider the removal of the maternity 'tail'.	Improved consistency in centrally held maternity stats, due to the existence of only one source for maternity data held centrally.	Removal of the maternity 'tail' is to be investigated
Assessment Unit (AU) Activity	Inconsistency in approaches to recording assessment activity across Wales.	A national review is being undertaken to consider an appropriate approach to the future recording and reporting of AU activity. For the purposes of financial costing only, an alternative approach to identify 'short stay' emergency activity (based on episode length) is being used by the WG Financial Information Strategy in the interim.	Availability of consistent data relating to assessment activity.	Not actively being reviewed.
Source of Admission / Discharge Destination	Inconsistency in values recorded in Source of Admission and Discharge Destination, particularly in records relating to transfers, causing difficulties in tracking patient journeys and deriving provider spell data.	A consultation with health boards in September 2013 revealed little appetite for changes in national definitions. Compliance with data quality standards continue to be monitored with issues being addressed on a case-by-case basis.	Improvements in the accuracy of these data would allow for stricter logic in scripts used to derive provider spells and greater accuracy in related analyses.	New values were implemented April 2019. Analysis has shown mixed uptake of new values and large use of retired values, issues remain.
Well babies	Data is being reported inconsistently across Wales.	Review and refine national standards.	Improved consistency in data held centrally.	National standards relating to all babies (well and unwell) are not currently being reviewed.

Issue	Impact	Proposed Resolution	Benefit	Status
Non consultant activity	Non consultant activity is not currently recorded in the dataset giving an incomplete picture of inpatient activity	Incorporate non-consultant activity into the APC data.	Give a clearer picture of activity carried out.	National Standards not currently under review.
Treatment Function Code	620 has been invalid since 2015/16 as a Treatment Function code. 4 of 7 health boards submit monthly data under this value.	Address TFC value set to enable capture of GP led data	Allow Organisations to record GP led data within the dataset using valid reference data.	Meetings taking place to resolve the issue.
Subsidiary ICD Diagnostic Code	Some health boards populating subsidiary on all records as 2nd value. Others leaving blank and skipping to 1st secondary.	To retire the data item Subsidiary ICD Diagnostic Code from the APC Data set. To add an additional secondary ICD field (13th Secondary (ICD))	Consistent reporting and easier comparison of coding data across Wales.	All health boards other than Cwm Taf providing the 1st secondary diagnosis code.
Real Time data	COVID-19 pandemic has reinforced the need for closer to real-time data when influencing decision making.	To update the APC reporting schedule to a weekly submission of data. To increase APC reporting scope to include unfinished episodes.	To capture the most up to date data available to keep as close to real time as possible	Implemented September 2020. Post implementation review to follow to gauge change in submissions.