



Sollis Commissioning Technical Overview

Johns Hopkins University's Adjusted Clinical Groups

Introduction

Adjusted Clinical Groups (ACGs):

- Analyse population health needs using inpatient, outpatient, A&E and primary care data. Individuals are assigned to an ACG; members of each ACG have a similar pattern of morbidity and resource consumption over the course of a given year.
- Provide forecasts of costs and financial risk for a given period.
- Identify patients likely to be high future resource users, and those who could benefit from more intensive (or other alternative) management. They therefore support a variety of admission avoidance programmes.

The Sollis Commissioning Suite

Sollis Commissioning is a patient and population management suite of tools which enable primary care clinicians to focus on improving patient health outcomes and more effectively utilise community care nurses and case managers.

In direct response to David Nicholson's challenge of doing more with less, Sollis Commissioning brings together primary and secondary care data, supporting efficient resource allocation and performance management. All reported care activity is fully-costed, encouraging the design of innovative, more cost effective care services.

Johns Hopkins University Adjusted Clinical Groups

Johns Hopkins University Adjusted Clinical Groups (JHU ACGs) is a population based case-mix system for identifying and measuring the morbidity burden of primary care patients, based on disease patterns, age and gender. It supports the emerging GP commissioning agenda by enabling clinical commissioners to find innovative ways of managing multi-morbidity patients, redesigning primary and community services and developing pathways to improve outcomes. Used extensively worldwide, the ACG grouper is highly peer reviewed, ensuring inherent quality in its outputs.

Sollis Commissioning

Technical Architecture

The Sollis Commissioning suite utilises a client-server architecture based on a series of industry standard components, comprising an SQL server machine and one or more web servers.

The server environment exploits standard Microsoft technology including:

- Windows Server 2008 R2 Standard (64 bit)
- SQL Server 2008 R2 (64 bit), including SQL Server Integration Services (SSIS) for back-end data management; SQL Server Analysis Services (SSAS) for On-Line Analytical Processing (OLAP) capabilities; and SQL Server Reporting Services (SSRS) for common reporting capabilities
- Microsoft Office: Word 2003 or higher, Excel 2010 (64 bit)
- .NET framework 3.5 SP1
- Some customers additionally deploy Sollis dashboard reports in a SharePoint 2010 presentation environment

SSRS requires SQL Server 2008 R2 Standard or Enterprise edition 64 bit (or 32 bit where the web servers are running a 32 bit version of Windows). The SQL server requires only a Standard licence, while web servers require Enterprise licences.

The Sollis implementation of these technologies is fully open. No Sollis products or data structures are hidden and customers with appropriate skills are encouraged to develop additional services on top of the Sollis platform. This may include additional reports and Hypercubes (see below), new data stores and feeds, and additional processing rules. Many Sollis customers have developed data feeds to other information systems and resources. Sollis provides documentation and training to support this model.

Sollis also provides specialist applications written in C#. These include the CMS and AIV solutions, the PBC framework and SP.Hypercube – a powerful, multi-functional OLAP client, which can analyse and present information on a wide range of data sources and types. These are fully documented components available to end users, with numerous configuration options.

All of these technologies and components are deployed under customer control

Interfacing tools are available to support a range of other software tools and data formats, from Microsoft Office products to business class analytical tools, and additional plug-ins can be developed by Sollis or our customers. All of these technologies and components are deployed under customer control, and can be used to support open, multi-system architectures.

The Sollis product suite exploits all required UK National and NHS data definitions and standards to ensure that information is readily transferable to similarly compliant information solutions.

The Sollis product suite is demonstrably scalable to readily support cluster and multi-commissioner environments, and a range of specialised tailoring options are available to support such contexts. This removes the need for multiple, disparate standalone databases.

Information Governance and Third Parties

Sollis has had an N3 connection, together with appropriate supporting information governance (IG) and security compliance policies, since 2006. Sollis reviews its security policies continually to ensure they meet the latest relevant NHS standards, including ensuring completion of the IG toolkit.

Sollis is ISO 9001:2008 certified. Adherence to the relevant legal frameworks forms part of our Quality Management System (QMS) compliance process.

Apollo Medical Systems is the only third party Sollis would involve in data handling. Apollo is mandated by the Department of Health for primary care data extraction relating to the performance of each practice against their Quality and Outcomes Framework (QOF) targets. They meet all prerequisites for conforming to NHS guidelines.

Where third parties are brought in for change management and training purposes, any local non-disclosure and non-divulgence requirements will be adhered to.

Primary Care Data Transmission

All patient data when transmitted to the PCT (or equivalent NHS organisation) during an Apollo extraction is encrypted using industry standard, NHS compliant encryption cyphers, and is transmitted over N3. Once the data is within the host organisation's Sollis data warehouse, it is deemed to be within a safe haven.

Patient level data is only visible to clinical users who have been set up with the sufficient, agreed access privileges. Where practice users take downloads from the reports containing patient information, it is expected that NHS guidelines are followed in the storage and transport of this data by the user.

No primary or secondary care data is held within servers hosted by Apollo or Sollis. Data is transmitted directly from practice to PCT or equivalent over an N3 connection (point-to-point transfer).

Secure User Access

All data importing takes place on the customer's network domain. The import process does not require Sollis to view patient identifiable data. Once the data is imported, users only have access to this data via the ACG reports, which are secured according to user configuration in Clarity. User access to reports is logged for audit purposes.

Practice users can view reports that show aggregated figures at the commissioning organisation or locality level. They can also drill-down to patient data only for the practice to which they are assigned. Practice users are assigned to a practice in the Sollis Commissioning web site.

Community Matrons, who may need to view patient information for several practices, can be assigned a Clarity user role allowing them access to multiple practices.

There are two levels of access to patient data, which can be set for different types of practice user: 'read' and 'read restricted'.

At the 'read' level the user can see patients but in reports their ID, age, gender and LSOA are hidden, and the ability to drill down to the patient clinical profile is removed. This might apply to Public Health Team members.

At the 'read restricted' level users can see the sensitive fields listed above, and can drill-down to the patient clinical profile. This would typically apply to GPs or Community Matrons.

Non-clinical commissioning organisation users are assigned to a commissioning organisation in the Clarity Security Manager (this is a software component of the Sollis solution, operated by local customer security administrators). Such users cannot see data at the locality or practice level, only at the PCT level. Furthermore they cannot view patient level data. They can filter results by LSOA, however they can see only view-aggregated results and there is no facility to drill-down to detail like in the practice user reports. While LSOA is spatial data, the current reports do not enable it to be used to say which practice the numbers belong to.

Johns Hopkins University: Adjusted Clinical Groups

The Predictive Model

The ACG System includes three types of predictive models, Dx-PM, Rx-PM and DxRx-PM, which differ according to the data inputs. The first is based on diagnostic codes (ICD-9-CM, ICD-10, Read, ICPC and so on), the second is built with NDC or ATC medication codes, and the third uses both diagnostic and medication code inputs.

Adjusted R-Squared (R²) Values

The conventional measure of model performance is the R² statistic. This statistic measures how well the model fits the data and has become a standard measure of performance, especially among underwriters and actuaries, who must price products across a range of populations. A value of 100% would indicate that the model explains all of the variance in the dependent variable of interest.

Global application of the ACG System has proven its validity and robustness under all types of healthcare systems. The illustration below shows R² values from several countries when using diagnoses to explain the variance in various resource measures in the concurrent year.

Country	Dependent Variables	Age, Gender	Age, Gender, ADGs	ACGs Alone
United States	Total Costs (Including Pharmacy)	13%	55%	37%
Canada-Manitoba	Ambulatory Costs	8%	50%	43%
Taiwan	Physician Costs	12%	52%	47%
	Dr. Visits	6%	58%	53%
Sweden	Primary Care Costs	11%	NA	38%
Spain	GP Visits	13%	59%	53%
United Kingdom	GP Visits	-	-	54%

The Sollis implementation of ACGs combines diagnostic and medication information. As the narrative explains, this gives the strongest predictive results.



The table below shows R² results when using diagnoses as well as medication based models to predict year two resource use.

Predictive Model	Total Costs	Pharmacy Costs	Physician Visits
Age, Sex	5%	7%	5%
Diagnostic Based Models			
Age, Sex ADG	17%	20%	25%
ACG	16%	17%	22%
Age, Sex, Charlson Disease Score	13%	17%	10%
Dx-PM	21%	29%	23%
Medication Based Models			
Age, Sex, Chronic Disease Score	16%	35%	12%
Rx-PM	19%	44%	17%
Combined Models			
DxRx-PM w/out Prior Costs	24%	46%	-
DxRx-PM with Prior Costs	26%	57%	-

Sensitivity and Positive Predictive Value

An increasingly important use of the ACG Suite of Predictive Models has been for high risk case identification. For this application, models are used more like a diagnostic test and performance is measured at how well true cases are identified and false positives ignored. The focus tends to be on two key indicators: sensitivity and positive predictive value.

The computational approach for these indicators is as follows:

- Sensitivity = $\frac{\text{True Cases Identified}}{\text{All True Cases In Population}}$
- Positive Predictive Value = $\frac{\text{True Cases Identified}}{\text{All Cases Identified}}$

Sensitivity is likely to be of greatest interest to epidemiologists and others who are focused on the health of the population since they are considering how well the 'test' captures all of the high risk individuals in a population. Therefore a value of 100 would indicate that all true cases in a population were identified.

Positive predictive value will likely be of greatest interest to clinicians/care managers who want to know the likelihood that a particular patient is actually high risk. Similarly, a value of 100 would indicate that all cases identified were in fact true cases.

Predictive Model (Predicting Total Cost)	Sensitivity and PPV*
Prior Cost	35
Age and Gender	20
Dx-PM w/o Prior Cost	33
Rx-PM w/o Prior Cost	34
DxRx-PM w/o Prior Cost	37

*Sensitivity and PPV are balanced in this example because the size of the group of the highest spenders matches the size of the highest risk group. Each group consists of 5% of the population. When one identifies a group at risk for being high spenders that is of the same size as those that will actually be classified as high spenders, PPV and Sensitivity are, by definition, the same. One would separate Sensitivity and PPV when the two group sizes are not equal.

C-Statistics

The C-Statistic is a measure of the probability that the model used will correctly identify true positives. A C-Statistic of 0.5 indicates that true cases are indistinguishable from false positives (or a model no better than chance). A C-Statistic of at least 0.8 is widely accepted as a threshold for good test performance. For models predicting total cost with a 5% cut point (that is, the top 5% of actual year two costs defines high risk), C-statistics are summarised in the table below.

Predictive Model (Predicting Total Cost)	C-Statistic
Dx Alone	.820
Dx + Prior Total Cost	.833
Rx Alone	.797
Rx + Prior Total Cost	.802
DxRx Alone	.830
DxRx + Prior Total Cost	.835

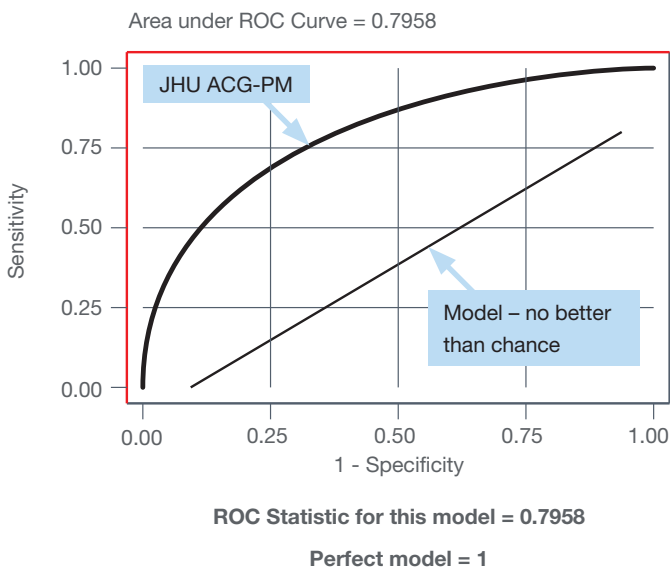
The following risk factors are used in the Johns Hopkins ACG Predictive Model (ACG-PM).

Risk Score	◀ Age
	◀ Gender
	◀ Complicated Pregnancy Marker
	◀ Pharmacy Use Marker (Optional)
	◀ Hospital Dominant Conditions
	◀ Selected Medical Conditions
	◀ Overall Disease Burden

ROC Curve

The Receiver Operating Characteristic (ROC) curve below compares the ACG predictive model with a hypothetical 'perfect model' and a model which is no better than chance. The current version of the ACG system shows a ROC score of just under 0.8 which is not bettered by any available in the UK.

A key benefit of ACGs is that JHU invest the licence fees back into research and development to improve the system. Thus the newer version of ACGs (version 9) will incorporate some prior utilisation data factors which further improve the performance.



Data and Risk Re-Calculation Frequency

In the majority of Sollis ACG implementations secondary care data is refreshed on a monthly cycle. Primary care data can be processed within the system anything between weekly and quarterly, depending on requirements. In large implementations (e.g. Strategic Health Authority-wide) the data feeds from GP practices are typically staggered, to help reduce potential processing bottlenecks. In most implementations of the ACG solution the primary care data is processed either monthly or quarterly.

Risk scores are typically recalculated weekly or monthly, depending upon the detailed scheduling of the various data feeds.

Recalibration

The model can be locally tailored to better reflect the distribution of Resource Utilisation Bands (RUBs). If the ACG system is serving a sufficiently large population, the allocation of RUBs can be localised to better represent the morbidity burden of the population being served.

Modest performance improvements may be attained when predictive models are recalibrated against a local population. The ACG software facilitates local calibration of predictive models with the export of the independent variables as model markers.

Firstly the ACG software calculates local weights (where cost data in input) to create rescaled weights. The software-supplied reference ACG-weights are supplied in two forms: un-scaled and rescaled. Un-scaled ACG-weights are simply the values of the reference ACG-weights applied to a population of interest. The mean value of the un-scaled ACG-weights provides a rudimentary profiling statistic. If the mean of the un-scaled ACG-weight is greater than 1.0, it indicates the rating population (the population to which the weights are being applied) is sicker than the reference population (the national reference database). If the mean is less than 1.0, it indicates the rating population is healthier. To ensure that costs in the system are not over or under-estimated, Sollis has also made available a rescaled or standardised ACG-weight that mathematically manipulates the un-scaled ACG-weight to have a mean of 1.0 in the local population.

If local cost data is available, the ACG software also calculates local ACG-weights. These local weights more accurately reflect local benefit levels and area practice patterns. In general it is recommended that the reference population (on which the weights are developed) should be as similar as possible to the assessment population to which the weights are applied.

Secondly, the software provides a report of the input variables used for calibration, or for users to create their own regression models, for the purposes of validation or to create additional predictive models based on local data.

Data Requirements

The three main inputs in terms of primary care are:

- **Patients**
A complete list of all the registered patients within the practice including date of birth and gender. The patient's NHS Number is used to link with data from secondary care. This is supplemented where necessary with other fields.
- **Events**
A full extract of all the Read coded events (per patient) held on the practice system for a period of 12 months. In addition to this, long-term conditions (LTCs) diagnoses recorded on the practice system going back as far as possible are also extracted.
- **Encounters**
An extract containing all of the different types of contacts that occurred in the past 12 months.

Primary and Secondary Care Data

Practices are not required to extract the data themselves. An automated process can be set to run out of hours. For practice systems where an automated process is not available or may be unsuitable, Sollis can offer a MiQUEST based solution.

The integrated system also incorporates secondary care (SUS) activity for the relevant practice's population to create a combined primary and secondary care data set. This contains both the Read coded information from the practice and any secondary care activity recorded by providers (with diagnoses coded to the ICD10 standard).

The primary care data extraction process is an out of hours, automated job that extracts the required data at the practice, encrypts it and sends it via N3 to the host PCT or commissioning organisation. There is no human intervention required for this process to run either for the first extraction or for the regular monthly or quarterly extractions. No patient data is held by either Apollo or Sollis at any point (not even in encrypted form).

Depending upon the agreed level of support built into the project, Sollis would either be responsible for updating the ACG data on a frequency agreed with the customer, or provide the means for the user to carry this out themselves as part of any routine processing they may already be doing.

Secondary care data is processed within the Sollis system. The customer is expected to download the relevant SUS Extract Mart from SUS on a regular basis and import it into the Sollis data warehouse as part of their routine processing. This is then used when processing the secondary care data as part of the ACG analysis.

Incomplete Information

When primary and secondary care data is imported into the Sollis data warehouse it is validated and verified to minimise the likelihood of missing data. Checks for missing data include:

- Checking the integrity of the import file structure:
 - Missing periods of activity
 - Gaps in data between extracts
- Checking for missing populations (you would expect a population spread across an age range)

The primary care extraction process will not extract the details of any patient who has requested that their records not be shared, and where their records have been Read coded accordingly.

Data Validation

The ACG grouper validates clinical data and produces a validation warnings report which includes:

- Patients older than 107 years
- Patients pregnant but not female
- Patients pregnant but not of child bearing age
- Indication of delivery but not of pregnancy
- Patients with 0 total cost but with diagnosis
- Patients with 0 pharmacy cost but with prescriptions
- Unmatched diagnosis codes (Read and ICD)

It validates, using regularly updated mapping files, clinical coding and reports any Read or ICD codes that have not been mapped to any specific condition in an un-matched diagnosis report.

The ACG grouper also reports details of missing drug costs, where no British National Formulary (BNF) mapping is available or no cost is available on the annual Prescription Cost Analysis (PCA) from the Information Centre.

Secondary care data validation is provided by the Sollis Commissioning software. Warnings and errors are also logged during the practice data loads.

Other validation includes:

- Checking for gaps in the data between consecutive extracts
- Checking for missing data between overlapping periods in the consecutive feeds
- Checking to ensure initial extracts are loaded prior to regular loads
- Checking for correct formatting and consistent header information of the files
- Checking for patients who have left and returned to ensure their old data is integrated

The next release of the Sollis ACG solution will incorporate further analysis and validation of the incoming data streams.

Patient population and range checks are based on rules and thresholds, such as date of birth being available for no less than 99% of the population.

Patient records which are available from the secondary care data sets but for which there are no matches found in the primary care data sets are flagged to support exception reporting.

Reporting

As part of a standard implementation, reports are presented through a web enabled reporting portal, which is delivered over a secure N3 connection and is fully compatible with web standard encryption protocols (<https://>) as long as the prerequisite certificates are in place.

Standard outputs are Microsoft Excel and PDF files from SQL Server Reporting Services (SSRS) reports, which are most useful where data is graphical. Other options are as XML, CSV, TIFF and web archive.

The implemented system enables direct access to the reporting tables and views, so users can query the raw data as required for ad hoc reporting. Users can create reports in SSRS, MDX cubes in SQL Analysis Services (SSAS), and Hypercubes. (SP.Hypercube can also work with MDX cubes, which is recommended.) The Sollis training programme includes details on the table schemas and a full data dictionary.

Further, the 'Patient Viewer' provides an integrated view of all of the care received by a patient, across all settings.

Any reports containing patient identifiable data are only visible to users with the appropriate access rights.

Flexible Reporting

Ad hoc reporting is provided as standard via a suite of Hypercubes. These On-Line Analytical Processing (OLAP) cubes allow the user to create and save reports at practice, patient and commissioning group levels.

Hypercubes can also provide further detailed analyses relating to:

- LSOA (for geographic analysis)
- Age
- Gender
- RUB (that is, risk factor: Non-Users, Healthy, Low, Moderate, High, Very High)
- Cost
- Locality (borough, commissioning group)
- Practice
- Patient (only available to practice users who have been configured to have appropriate privileges)

Users can also create their own custom cubes, or custom cubes may be developed for customers by Sollis.

Access to the Database and Raw ACG Extracts

The Sollis data warehouse is an open environment, giving users with appropriate user rights the ability to query the raw data as required. Some key fields available are:

- NHS number
(usually taken from any secondary care activity)
- Local patient ID
(the patient ID as recorded on the local practice system)
- Hospital patient ID
(for linking to other data from secondary care)
- Practice code
(six character code)
- GP code
(eight character GMP code)
- PCT code
(the three character host commissioner code)

These allow linking to other data sets outside of the ACG data sets at patient, practice and GP, and commissioning organisation level. Similarly, bespoke offloads can be created for use by a wide range of other reporting solutions.

The standard reporting outputs also contain a Lower Super Output Area (LSOA) for geographical population morbidity burden analysis or linking to GIS software.

Secondary care activity records provide the Unique Booking Reference Number (UBRN) for referrals and other records, enabling links to 'Choose and Book' data sets (to support referral management initiatives, for example).

Risk Reporting

Risk reporting enables the user to assess the morbidity burden of their population from a clinical perspective by focusing on:

- Cohorts of patients in specific risk groups (RUBs)
- Patients with specific long-term conditions
- Patients whose risk score has increased compared to their previous score
- Patients who have a high number of conditions likely to cause a need for hospitalisation

Risk scores are allocated per patient, and a patient is the lowest level of reporting available. This allows the reporting to be aggregated up into practice, commissioning group and PCT level.

Standard reporting facilitates:

- Benchmarking
- Risk analysis
- Statistical analysis
(measuring relative practice efficiency)
- Anonymised record level outputs

Other, locally specified reports could be developed.

Scalability

The solution is capable of supporting multiple PCTs. The Sollis Commissioning solution is a web based portal which is accessible via an N3 connection. Each of the PCTs (commissioning organisations) would have their own portal providing them with information across their population.

The solution is deployed in large data centre environments as well as single NHS organisations. It is deployed within an environment that has been scoped to support circa four million patients, across nine PCTs.

The infrastructure supporting the deployment (SQL and web server, N3 bandwidth and storage) would need to be scalable in line with the expected increase in data (processing and storage) and users accessing the system.

Sollis

20 Hook Road, Epsom
Surrey KT19 8TR

T: 01372 847 525
W: www.sollis.co.uk

