



# The Revised KC60 Statistical Return: Genitourinary Medicine Clinic Activity Dataset (GUMCAD)

---

## Technical guidance and specification for data extract from GUM clinics

28<sup>th</sup> February 2008

ISB Approved Version (13)

---

Prepared by:

**Dr Gwenda Hughes**

*Head of STI Section*

and

**Ms Geraldine Leong**

*Senior Scientist*

*Health Protection Agency Centre for Infections*

**Dr Immy Ahmed**

*President*

*British Association for Sexual Health and HIV*

on behalf of the Revised KC60 Steering Group  
and the BASHH Information Group

## CONTENTS

---

<b>BACKGROUND .....</b>	<b>3</b>
<b>AIM OF DOCUMENT.....</b>	<b>3</b>
<b>DATA EXTRACT SPECIFICATION .....</b>	<b>4</b>
Description .....	4
Time period .....	4
Frequency.....	4
Format .....	4
Transmission .....	4
Notes.....	4
<b>DATA CODING SPECIFICATION.....</b>	<b>7</b>
Figure 1. Format for the GUMCAD extract.....	7
Figure 2. Example of CSV format for the GUMCAD extract (for one row of data) .....	7
Figure 3. Coding specification for GUMCAD extract .....	8
<b>APPENDIX: KC60 CODES AND DESCRIPTIONS.....</b>	<b>14</b>

## **BACKGROUND**

---

In England, the monitoring of Primary Care Trust (PCT) Local Delivery Plans (LDPs) require data on patients attending GUM services by PCT of residence. Residence based rates of sexually transmitted infection (STI) diagnoses and testing are required to help monitor LDPs as well as to estimate the burden of infection and assess local STI screening practices. Currently, the only routinely available data on STI diagnosis and screening is compiled by data submitted to the HPA on the KC60 statistical return, but these returns collect no information on patient area of residence. Although some residence-based data have been extracted locally, there is currently no routine extraction of residence-based information across England on patients attending GUM services.

BASHH and the HPA, in consultation with the Department of Health, have collaborated to specify the minimum data needed to meet these requirements. The data specified will be compliant with the proposed Common Data Set for Sexual Health. Collection and reporting of the data items should be possible at most GUM clinics and it is intended that the implementation should begin in April 2008. This document specifies the minimum data set for the implementation.

## **AIM OF DOCUMENT**

---

To provide specifications for the revised KC60 data extract, known as the Genitourinary Medicine Clinic Activity Dataset (GUMCAD), from genitourinary medicine clinic software.

## DATA EXTRACT SPECIFICATION

---

### Description

Each GUM clinic will be required to generate a quarterly data extract of all patient attendances and associated diagnoses (coded using KC60 codes) at GUM clinics by patient PCT of residence, Lower Super Output Area of residence, gender, age, sexual orientation, ethnicity and country of birth for emailing to PCTs, SHAs and the HPA. It is proposed that the GUMCAD extract, once validated and with some modification, will eventually replace the current KC60 statistical return. All specified fields will be mandatory.

### Time period

The extract will cover one calendar quarter.

### Frequency

Reports will be run quarterly, 6 weeks after the end of the quarter.

### Format

Data returned should be formatted into a single comma delimited CSV file or in Microsoft Excel format. The format of the CSV/Excel file is presented in figure 1. An example of the field content is also shown and is used to illustrate how the data should appear in the CSV file (figure 2). The field content and field order should be identical for files in Excel format.

### Transmission

Electronic files will be transmitted to the HPA. Where possible, data should be submitted to the HPA through the secure document gateway in the HPA website. This gateway enables organisations to distribute any type of files to previously identified users in a secure manner across the Internet. The document gateway can be found at: <https://www.hpa.org.uk/login.spl> .

Connection to the gateway requires a login account name and password, which will be available from the project administrator at the Centre for Infections. The browser supports the Secure Sockets Layers (SSL) method of communication and passwords are changed every 3 months.

### Notes

#### 1) *Definition of first attendances for GUMCAD*

First attendances refer to all new or re-book face-to-face attendances at the start of a genitourinary episode. All attendances by an individual patient less than 26 weeks after their last attendance will be routinely classified as a follow-up attendance by clinic software, unless manually altered on the system at the clinic. All attendances by an individual patient 26 weeks or more after their last attendance will be routinely classified as a re-book attendance by clinic

software, unless manually altered on the system at the clinic. When a clinician decides that an episode has been completed within the 26 week interval, clinic staff should manually close the episode using the clinic software. First attendances will be identified in the extract in the 'First\_Attendance' field (which records attendance type). The majority of first attendances will have an associated KC60 code when transmitted. However, the KC60 field may occasionally be blank when transmitted where there has been a delay in diagnosis. All first attendances must eventually have an associated KC60 code.

## 2) *Definition of subsequent attendances*

Subsequent attendances refer to all follow-up attendances within a genitourinary episode. Dates of subsequent attendances will be included in the GUMCAD extract to allow generation of part C of the current KC60 statistical return. Subsequent attendances will be identified in the extract in the 'First\_Attendance' field (which records attendance type). The majority of subsequent attendances should not have a KC60 code as they are related to previously recorded diagnoses (they are not new diagnoses) and for these visits the 'KC60' field should remain blank. There are two exceptions (1). Where a patient receives a new diagnosis during a follow up attendance for an existing condition. (2). Where a new KC60 code is appropriate for a change in disease status e.g. E1B, E2B and E3B. All follow up attendances (regardless of whether they have a KC60 code or not) should be recorded as such in the 'First Attendance' field.

## 3) *Allocation of missing PCT of residence*

Records where the patient's postcode has not been provided to generate PCT of residence should be allocated to 'not known' and coded "Q99". Postcodes outside of England (overseas visitors, Wales, Scotland or Northern Ireland) should be allocated to 'not applicable' and coded as "X98".

## 4) *Allocation of missing LSOA of residence*

There is a Lower Layer Super Output Area for each postcode in England and Wales. Postcodes in Scotland, Northern Ireland, Channel Islands or Isle of Man should be coded "Z99999999". Records where the patient's postcode has not been provided to generate LSOA of residence should be allocated to 'not known' and coded "X99999999". Postcodes outside the United Kingdom should be allocated to 'not applicable' and coded "X99999998".

## 5) *Allocation of changing PCT/LSOA of residence*

Reports should use the latest PCT/LSOA of residence available.

## 6) *Allocation of changing patient details*

If a patient's demographic details (e.g. ethnic group, sexual orientation etc.) are changed during a given month, the most recent entry should be allocated to the record being transmitted.

### 7) ***Recording sexual orientation***

Clinical staff taking sexual histories are advised to enter accurate information during the clinical consultation or as soon as possible thereafter. Clinic software systems should avoid defaulting patients' sexual orientation to 'heterosexual', rather software should default to 'Not known'. To improve the quality of sexual orientation information recorded, software should prevent completion of data entry until sexual orientation has been recorded.

### 8) ***Recording and transmission of KC60 codes***

KC60 codes should not be grouped as some currently are on the KC60 statistical return. Instead, each individual code should be transmitted separately. For example, 'C4A,C4C' should be transmitted separately as either C4A or C4C, as appropriate (see Appendix).

### 9) ***Coding and formatting***

The coding and formatting of the data items required are presented in figure 1. The full list of KC60 codes for the GUMCAD extract is provided in the appendix. All fields specified are compliant with the proposed Common Data Set for Sexual Health ([www.cdssexualhealth.org.uk](http://www.cdssexualhealth.org.uk)) and are defined using the NHS data dictionary where possible.

## DATA CODING SPECIFICATION

**Figure 1. Format for the GUMCAD extract**

Position *	Field Name	Description	NHS Data Dictionary Data Element	Variable Length <sup>‡</sup>	Example <sup>‡</sup>
1	ClinicID	Clinic ID code	<a href="#">SITE CODE (OF TREATMENT)</a>	AN(5)	RCC25
2	PatientID	Local patient identifier number	<a href="#">LOCAL PATIENT IDENTIFIER</a>	AN(10)	PAT123
3	KC60	KC60 code	<a href="#">GENITOURINARY EPISODE TYPE</a>	AN(4)	C10A
4	Gender	Gender	<a href="#">PERSON GENDER CURRENT</a>	N(1)	1
5	Age	Age at attendance date in years	AGE AT ATTENDANCE DATE	N(3)	16
6	Sex_Ori	Sexual orientation	SEXUAL ORIENTATION (CURRENT)	N(1)	1
7	Ethnicity	Patient's ethnic category	<a href="#">ETHNIC CATEGORY</a>	AN(2)	A0
8	Country_Birth	Patient's country of birth	COUNTRY CODE (BIRTH)	A(3)	GBR
9	PCT	PCT of residence code	<a href="#">ORGANISATION CODE (PCT OF RESIDENCE)</a>	AN(3)	5K9
10	LSOA	Lower Super Output Area of residence code	LOWER LAYER SUPER OUTPUT AREA (RESIDENCE)	AN(9)	E01000001
11	First_Attendance	Attendance type	<a href="#">FIRST ATTENDANCE</a>	N(1)	1
12	AttendanceDate	Date of attendance	<a href="#">ATTENDANCE DATE</a>	N(10) CCYY-MM-DD	2007-10-31

\*Refers to the horizontal position of the field within CSV format

<sup>‡</sup>AN = Alpha-numeric, N = Numeric, A = Character. Number in brackets denotes the string length.

<sup>‡</sup>Example of field content, also used to illustrate extract format expected (see figure 2)

**Figure 2. Example of CSV format for the GUMCAD extract (for one row of data)**

```
ClinicID, PatientID, KC60, Gender, Age, Sex_Ori, Ethnicity, Country_Birth, PCT, LSOA, First_Attendance, AttendanceDate
RCC25, PAT123, C10A, 1, 16, 1, A0, GBR, 5K9, E01000001, 1, 2007-10-31
```

**Figure 3. Coding specification for GUMCAD extract**

Data item	NHS data dictionary name	Definition/comments
Clinic ID code (code to identify clinic/service provider)	<a href="#">SITE CODE (OF TREATMENT)</a>	<b>Format/length:</b> an5 An identifier for a <a href="#">CLINIC OR FACILITY</a> .
Patient ID number	<a href="#">LOCAL PATIENT IDENTIFIER</a>	<b>Format/length:</b> an10  <b>Note:</b> This is a number used to identify a <a href="#">PATIENT</a> uniquely within a <a href="#">Health Care Provider</a> . It may be different from the Patient's case note number and may be assigned automatically by the computer system.
KC60 code	<a href="#">GENITOURINARY EPISODE TYPE</a>	<b>Format/length:</b> an4  <u>National Codes:</u> The national KC60 codes and their definitions are given in the appendix.  <b>Notes:</b> KC60 records the <a href="#">FIRST ATTENDANCE</a> for the diagnosis and/or treatment of an infection or disease, during a <a href="#">Genitourinary Episode</a> . This field is usually only completed for the first attendance relating to a genitourinary episode – the majority of follow-up attendances should not have a KC60 code and for these visits this field should remain blank. (The main exceptions will be for codes E1B, E2B and E3B.)  A <a href="#">FIRST ATTENDANCE</a> is an indication of whether the face to face contact is the first occasion on which a <a href="#">PATIENT</a> is seen by any staff group of the Health Care Provider. It begins an episode of care which comprises all the contacts made by the relevant staff of the Health Care Provider regardless of changes of <a href="#">LOCATION</a> .  A genitourinary episode is a period of time during which a <a href="#">PATIENT</a> attends a <a href="#">Consultant Clinic</a> or a <a href="#">Nurse Clinic</a> for a genitourinary problem. Each episode will be for one <a href="#">GENITOURINARY EPISODE TYPE</a> . It is therefore possible for a <a href="#">PATIENT</a> to have concurrent Genitourinary Episodes. A <a href="#">PATIENT</a> may also have more than one Genitourinary Episode for the same <a href="#">GENITOURINARY EPISODE TYPE</a> over time. An episode is terminated either by a <a href="#">PATIENT</a> being formally discharged or not being in face-to-face contact with the service for at least six months
Gender	<a href="#">PERSON GENDER CURRENT</a>	<b>Format/length:</b> n1 <u>National Codes:</u> <b>0</b> Not Known - means that the gender of the person has not been recorded. <b>1</b> Male <b>2</b> Female <b>9</b> Not Specified – means indeterminate, i.e. unable to be classified as either male or female <b>References:</b> UK Government Data Standards Catalogue (GDSC), Version 2.0, Agreed 11.09.03. GDSC: <a href="http://www.govtalk.gov.uk/gdsc/html/default.htm">http://www.govtalk.gov.uk/gdsc/html/default.htm</a>

Data item	NHS data dictionary name	Definition/comments
		<p><a href="#">PERSON GENDER CURRENT</a> is the same as <a href="#">PERSON GENDER CODE</a> where the <a href="#">PERSON GENDER TYPE</a> equals '02 - Person Gender Current'.</p> <p>The e-GIF standard <a href="#">PERSON GENDER CURRENT</a> should be used for all new and developing systems and for XML messages. For existing CDS EDIFACT messages however, <a href="#">SEX</a> should be used.</p>
Age	<b>AGE AT ATTENDANCE DATE</b>	<p><b>Format/length:</b> n3</p> <p>This is usually derived as the number of completed years between the <a href="#">PERSON BIRTH DATE</a> of the <a href="#">PATIENT</a> and the <a href="#">ATTENDANCE DATE</a> . However, age can be manually entered in the absence of patient date of birth. Not known = 999, i.e. date of birth not known and age cannot be estimated</p>
Sexual orientation (males & females)	<b>SEXUAL ORIENTATION (CURRENT)</b>	<p><b>Format/length:</b> n1</p> <p>See <a href="#">SEXUAL ORIENTATION CODE</a> for the National Codes</p> <p>The current <a href="#">SEXUAL ORIENTATION</a> of a <a href="#">PERSON</a></p> <p><u>National Codes:</u></p> <p><b>1</b> Heterosexual  <b>2</b> Homosexual  <b>3</b> Bi Sexual  <b>9</b> Unknown</p> <p>The <a href="#">SEXUAL ORIENTATION</a> of a <a href="#">PATIENT</a></p>
Ethnicity	<b><a href="#">ETHNIC CATEGORY</a></b>	<p><b>Format/length:</b> an2</p> <p>The ethnicity of a <a href="#">PERSON</a>, as specified by the PERSON.</p> <p><u>National Codes:</u></p> <p>White  <b>A</b> British  <b>B</b> Irish  <b>C</b> Any other White background</p> <p>Mixed  <b>D</b> White and Black Caribbean  <b>E</b> White and Black African  <b>F</b> White and Asian  <b>G</b> Any other mixed background</p> <p>Asian or Asian British  <b>H</b> Indian</p>

Data item	NHS data dictionary name	Definition/comments
		<p> <b>J</b> Pakistani  <b>K</b> Bangladeshi  <b>L</b> Any other Asian background             Black or Black British  <b>M</b> Caribbean  <b>N</b> African  <b>P</b> Any other Black background             Other Ethnic Groups  <b>R</b> Chinese  <b>S</b> Any other ethnic group   <b>Z</b> Not stated   <b>Note:</b> <a href="#">ETHNIC CATEGORY</a> is the classification used for the 2001 census, replacing <a href="#">ETHNIC GROUP</a> in the flows through the NHS-wide Clearing Service.         </p>

Data item	NHS data dictionary name	Definition/comments
Country of birth	<b>COUNTRY CODE (BIRTH)</b>	<p><b>Format/length:</b> A(3)</p> <p>This is the country where the <a href="#">PATIENT</a> was born.</p> <p><b>COUNTRY CODE (BIRTH) is the same attribute <a href="#">COUNTRY CODE</a></b></p> <p>Refer to the ISO 3166-1 standard for actual list of alphabetic codes and countries. The alphabetic code to be used is the 3-char alphabetic code available on the International Organisation for Standardisation website <a href="http://www.iso.org/iso/home.htm">http://www.iso.org/iso/home.htm</a></p> <p><b>Note:</b> The 2-char alphabetic code <b>must not be used.</b></p> <p>Refer to the ISO 3166-1 standard for actual list of alphabetic codes and countries. Where country of birth is unknown please record this as ZZZ</p> <p>Notes: UK Government Data Standards Catalogue (GDSC), Version 2.0, Agreed 01.01.02. GDSC: <a href="http://www.govtalk.gov.uk/gdsc/html/default.htm">http://www.govtalk.gov.uk/gdsc/html/default.htm</a></p> <p>Max 3 Characters</p> <p><b>Reference:</b> <a href="http://www.iso.org/iso/home.htm">http://www.iso.org/iso/home.htm</a></p>
PCT of residence code	<b><a href="#">ORGANISATION CODE (PCT OF RESIDENCE)</a></b>	<p><b>Format/length:</b> an3</p> <p>The <a href="#">ORGANISATION CODE (PCT OF RESIDENCE)</a> is the same as the attribute <a href="#">ORGANISATION CODE</a> where the <a href="#">ORGANISATION TYPE</a> is national code PT - <i>Primary Care Trust</i>.</p> <p>This is the <a href="#">ORGANISATION CODE</a> of the Primary Care Trust derived from the <a href="#">PATIENT's POSTCODE OF USUAL ADDRESS</a>.</p> <p>For the purposes of sending Commissioning Data Set (CDS) messages to the Secondary Uses Service (regardless of how local systems hold the data) it is essential at present to continue using a 3 character field, using the first 3 characters of the ORGANISATION CODE (PCT OF RESIDENCE) and following the same update rules relating to Prime Recipient as are currently in place. This is necessary primarily to preserve the integrity of the current NHS CDS message and the Prime Recipient which is derived from the ORGANISATION CODE (PCT OF RESIDENCE).</p> <p>The National Administrative Codes Service (NACS) provides postcode files which link postcodes to the Primary Care Trust. See NHS Postcode Directory.</p> <p>Records where the patient's postcode has not been provided to generate PCT of residence should be allocated to 'not known' and coded "Q99". Postcodes outside of England (overseas visitors, Wales, Scotland or Northern Ireland) should be allocated to 'not applicable' and coded as "X98".</p>

Data item	NHS data dictionary name	Definition/comments
LSOA of residence code	<b>LOWER LAYER SUPER OUTPUT AREA (RESIDENCE)</b>	<p><b>Notes:</b> <a href="#">PCT OF RESIDENCE</a> is the same as the attribute <a href="#">ORGANISATION CODE</a>. See <a href="#">Primary Care Trust (PCT)</a> for the definitions of this <a href="#">ORGANISATION</a>.</p> <p><b>Format/length:</b> an9</p> <p>Notes: the Lower Layer Super Output Area for where the <a href="#">PATIENT</a> is resident. This is the <a href="#">GEOGRAPHIC AREA CODE</a> where the <a href="#">GEOGRAPHIC AREA TYPE</a> is classification Lower Layer Super Output Area.</p> <p>There is a Lower Layer Super Output Area for each postcode in England and Wales. Postcodes in Scotland, Northern Ireland, Channel Islands or Isle of Man should be coded “Z99999999”. Records where the patient’s postcode has not been provided to generate LSOA of residence should be allocated to ‘not known’ and coded “X99999999”. Postcodes outside the United Kingdom should be allocated to ‘not applicable’ and coded “X99999998”.</p> <p>Lower Layer Super Output Areas are a geographic hierarchy designed to improve the reporting of small area statistics. Lower Layer Super Output Areas are built from groups of contiguous Output Areas and have been automatically generated to be as consistent in population size as possible, and typically contain from four to six Output Areas.</p> <p>List of English Lower Super Output Area codes are available at  <a href="http://www.statistics.gov.uk/geography/soa.asp">http://www.statistics.gov.uk/geography/soa.asp</a>  <a href="http://neighbourhood.statistics.gov.uk/dissemination/Info.do?page=SOAConstitutions.htm">http://neighbourhood.statistics.gov.uk/dissemination/Info.do?page=SOAConstitutions.htm</a></p>
Attendance type	<a href="#"><b>FIRST ATTENDANCE.</b></a>	<p><b>Format/length:</b> n1</p> <p>The National Codes for ‘FIRST ATTENDANCE’ in the NHS Data Dictionary are:</p> <ul style="list-style-type: none"> <li>1 First attendance face to face</li> <li>2 Follow-up attendance face to face</li> <li>3 First telephone or telemedicine consultation</li> <li>4 Follow up telephone or telemedicine consultation</li> </ul> <p><b>Notes:</b>  This indicates whether a patient is making a first or follow-up attendance. For Genitourinary Clinic Attendances a <a href="#">FIRST ATTENDANCE</a> is the first in a series, or the only attendance by a <a href="#">PERSON</a> at a Consultant Clinic.</p> <p>Re-book should still be collected locally. It is used to identify frequently repeat attendees recurring infections which are of significance in managing individual patient care and for protecting public health.</p>

Data item	NHS data dictionary name	Definition/comments
Date of attendance	<a href="#">ATTENDANCE DATE</a>	<p><b>Format/length:</b> n10 – cyy-mm-dd</p> <p>ATTENDANCE DATE is the same as attribute <a href="#">ACTIVITY DATE</a> of <a href="#">ACTIVITY DATE TIME</a> where the <a href="#">ACTIVITY DATE TIME TYPE</a> is National Code 33 'Attendance Date'</p> <p>UK Government Data Standards Catalogue (GDSC), Version 1.0, Agreed 01.01.02. GDSC:  <a href="http://www.govtalk.gov.uk/gdsc/html/default.htm">http://www.govtalk.gov.uk/gdsc/html/default.htm</a></p> <p>This is the e-GIF standard that should be used for all new and developing systems and for XML messages</p>

## APPENDIX: KC60 CODES AND DESCRIPTIONS

Code	Condition/episode
A1	Primary infectious syphilis
A2	Secondary infectious syphilis
A3	Early Latent syphilis (first 2 years)
A4	Other acquired syphilis
A5	Syphilis of the nervous system (neurosyphilis)
A6	Late latent syphilis
A7	Congenital syphilis, aged under 2
A8	Congenital syphilis, aged 2 or over
A9	Epidemiological treatment of suspected syphilis
B1	Uncomplicated gonorrhoea - genital
B2	Pre-pubertal gonorrhoea
B3	Gonococcal ophthalmia neonatorum
B4	Epidemiological treatment of suspected gonorrhoea
B5	Complicated gonococcal infection – including pelvic inflammatory disease (PID) and epididymitis
C1	Chancroid
C2	Lymphogranuloma venerum
C3	Donovanosis
C4A	Uncomplicated Chlamydial infection
C4C	Pre-pubertal Chlamydial infection
C4B	Complicated Chlamydial infection – including PID and epididymitis
C4D	Chlamydia ophthalmia neonatorum
C4E	Epidemiological treatment of suspected Chlamydia
C4H	Uncomplicated non-gonococcal/non-specific urethritis in males or treatment of mucopurulent cervicitis in females
C4I	Epidemiological treatment of non-specific genital infection (NSGI)
C5	Complicated infection (non-chlamydial/non-gonococcal) – including PID and epididymitis
C6A	Trichomoniasis
C6B	Anaerobic/Bacterial vaginosis & anaerobic balanitis
C6C	Other vaginosis/vaginitis/balanitis
C7A	Anogenital candidosis
C7B	Epidemiological treatment of C6 & C7
C8	Scabies
C9	Pediculosis pubis
C10A	Anogenital herpes simplex: first attack
C10B	Anogenital herpes simplex: recurrence
C11A	Anogenital warts: first attack
C11B	Anogenital warts: recurrence
C11C	Anogenital warts: re-registered cases
C12	Molluscum contagiosum
C13A	Viral hepatitis B (Hepatitis B surface antigen positive): first diagnosis**
C13B	**number of which were acute viral hepatitis B
C13C	Viral hepatitis B: subsequent presentation
C14	Viral hepatitis C: first diagnosis
D2A	Urinary tract infection
D2B	Other conditions requiring treatment at GUM clinic
E1A	New human immunodeficiency virus (HIV) diagnosis: asymptomatic
E2A	New HIV diagnosis: symptomatic (not AIDS)
E1B	Subsequent HIV presentation (not AIDS)
E2B	Subsequent HIV presentation (not AIDS)
E3A1	AIDS: first presentation - new HIV diagnosis
E3A2	AIDS: first presentation - HIV diagnosed previously
E3B	AIDS: subsequent presentation

P4A	Cervical cytology: minor abnormality
P4B	Cervical cytology: major abnormality
S1	Sexual health screen (no HIV antibody test)
S2	HIV antibody test and sexual health screen
P1A	HIV antibody test (no sexual health screen)
P1B	HIV antibody test offered and refused
P2	Hepatitis B vaccination (1 <sup>st</sup> dose only)
P3	Contraception (excluding condom provision)
D3	Other episodes not requiring treatment