



McDaid, Lisa, Docherty, Stevie, and Winter, Andrew (2013) *Occasional Paper No. 24: Review of the National Sexual Health IT System (NaSH) in Scotland: the potential for sexual health research*. Other. MRC/CSO Social and Public Health Sciences, Glasgow.

Copyright © 2013 University of Glasgow

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

Content must not be changed in any way or reproduced in any format or medium without the formal permission of the copyright holder(s)

When referring to this work, full bibliographic details must be given

<http://eprints.gla.ac.uk/90011/>

Deposited on: 03 February 2014

**Review of the National Sexual Health  
IT System (NaSH) in Scotland:  
The potential for sexual health research**

**Lisa McDaid<sup>1</sup>, Stevie Docherty<sup>2</sup>, Andrew Winter<sup>3</sup>**

**1 MRC/CSO Social and Public Health Sciences Unit, University of Glasgow**

**2 College of Social Sciences, University of Glasgow**

**3 Sandyford Sexual Health Services, NHS Greater Glasgow and Clyde**



## Table of Contents

	<b>Page</b>
<i>Table of Contents</i>	<i>i</i>
<b>Abstract</b>	<b>3</b>
<b>Executive summary</b>	<b>4</b>
<b>1. Introduction</b>	<b>10</b>
<b>2. Methods</b>	<b>12</b>
<b>2.1 Policy and development review</b>	<b>12</b>
<b>2.2 Review of use of EPR systems in other sexual health clinics</b>	<b>13</b>
<b>2.3 Review of general methodological issues</b>	<b>14</b>
<b>3. Results and Discussion</b>	<b>16</b>
<b>3.1 What data are collected by NaSH, how comprehensive is the system, and what measures of data quality are available?</b>	<b>16</b>
<b>3.2 Data collection in NaSH</b>	<b>16</b>
<b>3.3 Anonymised data view</b>	<b>20</b>
<b>3.4 Data Linkage</b>	<b>21</b>
<b>3.5 ISD minimum dataset</b>	<b>22</b>
<b>3.6 NaSH data completion and completeness</b>	<b>24</b>
<b>3.7 Measures of data quality</b>	<b>27</b>
<b>3.8 What would be the key issues to address in routine use of NaSH to better yield benefits for sexual health research?</b>	<b>28</b>
<b>3.9 Data collection and completeness</b>	<b>28</b>
<b>3.10 Data storage</b>	<b>30</b>
<b>3.11 Data retrieval</b>	<b>31</b>
<b>3.12 Research governance</b>	<b>32</b>
<b>4. Conclusions and recommendations</b>	<b>34</b>
<b>4.1 Continuation of NaSH</b>	<b>34</b>
<b>4.2 Recording willingness to be contacted for research</b>	<b>34</b>
<b>4.3 Data quality</b>	<b>35</b>
<b>4.4 Computer assisted completion / self-completion</b>	<b>35</b>
<b>4.5 Data warehousing</b>	<b>35</b>
<b>4.6 Anonymisation process and relevance of data items</b>	<b>36</b>
<b>5. Acknowledgements</b>	<b>37</b>
<b>6. References</b>	<b>38</b>

<b>7. Appendices</b>	<b>41</b>
<b>Appendix 1: Clinic and parent department or hospital websites (sites accessed for data collection between November 2011 and May 2012; all links live at 06 September 2012)</b>	<b>41</b>
<b>Appendix 2: Year of system implementation, data collection methods, data collected, quality assurance and monitoring across clinics</b>	<b>42</b>
<b>Appendix 3: Example of relevant data items that can be updated at subsequent visits</b>	<b>43</b>

## Abstract

**Background:** Specialist sexual health settings present specific challenges to electronic record systems, including enhanced requirements from patients for discretion, anonymity, and rapid turn around. A number of international sexual health settings have pioneered bespoke electronic patient record (EPR) systems, and the National Sexual Health (NaSH) electronic clinical record system has been rolled out across specialist sexual health services in Scotland. The data collected could present a rich resource and here we discuss the key issues to address in the routine use of NaSH data for sexual health research. The intended audience for this report is primarily those in Scotland who are using, responsible for, or have the potential need to access, NaSH data, but the report will have wider relevance to those interested in sexual e-health.

**Methods:** Scoping review in three stages: a policy review of NaSH documentation; a review of EPR issues reported by an international selection of clinics known to be using computerized clinical systems; and a review of more general methodological issues related to the use of EPR.

**Results:** NaSH entails a data set with over 700,000 patients and more than 300,000 attendances recorded annually. Data include medical, family and sexual history, reproductive health and contraception, social and lifestyle factors, test requests/results, patient actions/recalls, prescriptions, symptoms, physical examination details, partner notification, and referrals. NaSH allows patient-centred choice of whether to use an anonymous identifier or CHI number, which could facilitate record linkage. Key issues in the use of the data are: data collection and completeness; storage and retrieval; and research governance. An anonymised data view has been created, but not all NHS boards complete all data fields, and even though minimum input criteria have been established, use of NaSH in real time has been problematic and variable. The data view only reflects current, visible data, and while episode-based data remain true, lifetime sexuality and smoking status, for example, can change over time, and the 'original' or preceding data are written over; precluding any longitudinal analysis. Similarly, longer term retention of NaSH data and availability to researchers out with the NHS are issues that have yet to be addressed.

**Conclusions:** Interrogating NaSH would enable research to make better use of existing sexual health data in Scotland, be cheaper than initiating large-scale surveys, and give access to high-risk populations, but would need to address conflict between the need for comprehensive and complete data for research purposes and the need for a routine clinical system to function in a routine way. Concerns over data collection, storage and retention should be considered within the context of the wider public health and research benefit.

## Executive summary

Specialist sexual health services present specific challenges to electronic patient record (EPR) systems implementation. The Scottish National Sexual Health System (NaSH), a centralised web-based electronic clinical record system, went live in March 2008 and is now in use across all specialist sexual health settings in Scotland. Interrogating NaSH would enable researchers to make better use of existing sexual health data in Scotland, be cheaper than initiating large-scale surveys, give access to high-risk hard-to-reach populations, and provide the means to respond to immediate research questions. However, uptake of the electronic record varies across Scotland, and the data are recorded as part of a routine face-to-face consultation for clinical purposes rather than as part of an academic research study. Accessing NaSH-derived data and linkage to other health or social care data sets will also raise ethical issues due to the highly sensitive nature of the data recorded. In this review, we assess the data collection and completeness, storage and retrieval, and research governance issues involved in accessing and using NaSH data for sexual health research in Scotland.

### Research questions

1. What data are collected by NaSH, how comprehensive is the system, and what measures of data quality are available?
2. What would be the key issues to address in routine use of NaSH data to better yield benefits for sexual health research?

### Methods

Scoping review in three stages: a policy review of NaSH documentation; a review of EPR issues reported by an international selection of clinics known to be using computerised clinical systems; and a review of more general methodological issues related to the use of EPR.

### Results

#### *NaSH use*

Health in Scotland is delivered by 14 geographically-located health boards along with various special boards. Sexual health services in each board area are led by a Lead Clinician in Sexual Health, usually an accredited specialist in Genitourinary Medicine or Sexual and Reproductive Health. All eleven mainland boards in Scotland now use NaSH as their primary record system for specialist sexual health. NaSH now holds sexual health records of over 700,000 patients, with around 300,000 attendances recorded annually across eleven mainland health boards in around 200 locations. Data forms include medical, family and sexual history, reproductive health and contraception, social history, test requests/results, patient actions/recalls, prescriptions, symptoms, physical examination details, partner notification, and referrals. All mainland boards in Scotland now use NaSH as their primary record system, but the uptake of full EPR remains variable with some clinics still relying on paper proformas. NaSH allows patient-centred choice of whether to use an anonymous identifier or NHS Community Health Index 'CHI' number (Scotland's version of the NHS number), which could facilitate record linkage.

## Data Completeness

In 2011, Information Services Division (ISD), part of NHS National Services Scotland, one of the special NHS boards, specified minimum input criteria for NaSH. However, a national overview of data completeness is still lacking, as at June 2013, due to technical difficulties in ISD accessing and analysing the national anonymised NaSH data view. This is essential to determine the utility or otherwise of using routinely collected NaSH data for sexual health research. Data completeness in NaSH may be prejudiced by multiple interlinked issues, including:

- **Local practicalities:** Use of NaSH in real time has been problematic in some locations due to connection speed, and the design of the system with multiple forms and data items. If the system is down, data are either omitted or entered post-hoc, often to a limited extent.
- **Minimum data set:** NaSH has data fields for social and lifestyle risk factors, but the majority of these (except for sex, age, postcode and ethnicity) are not currently included in the ISD minimum dataset so boards are under no obligation to record them. Only those items on the ISD minimum dataset have so far been included in ISD's Business Objects universe (a custom anonymised view of NaSH data).
- **Data items missing from specification:** Even with the comprehensive system specification some data items of social health interest may have been overlooked. For example, record of interventions to reduce harmful drinking, or a space to record actual AUDIT or FAST scores for alcohol screening are still to be added to the system. This problem was repeatedly cited as a limitation in studies we reviewed from clinics around the world.
- **Differences between NHS Boards:** Some NHS Boards have chosen to retain partial paper or self-completed records and not to complete certain elements of the EPR. Although the proformas might be scanned on to facilitate clinical working, the content cannot be captured for analysis. Some Boards are only recording positive 'hits', i.e. not negative responses, on the form. This means when reporting on specific items at a national level one needs to be aware of regional differences in data quality and collection.
- **Human factors:** These are routinely collected, not research, data and, even if local policy dictates collection of a data item, this may not be completed on the day due to human factors, staff or patient fatigue or distress, oversight or unwillingness or discomfort addressing certain areas of the sexual risk history.
- **Targeted completion:** We found examples of variable data quality from NHS Greater Glasgow & Clyde, specifically that young people aged under 20 had fuller recording of social risk, and men of all ages were more likely than women to have their lifetime sexuality recorded. Attempts to use such data at a national level would require further analysis to assess whether they are part of a wider pattern, or are indicative of a wider tendency to record risk most completely for patients who are already perceived to be at risk: targeted completion, in other words.
- **Functional non-completion:** In day-to-day clinical practice, a clinician may view it as unnecessary to go over recent sexual or social risk history with a return patient attending for a routine contraceptive implant removal and re-insertion. However, the data items may

be of great interest to researchers. This has the potential, like targeted completion, to bias results obtained.

- **Mandatory data items:** There are few mandatory data items to avoid users getting 'stuck' during clinical use. This is very different to systems designed for research data entry that generally have a high level of mandatory data fields to ensure no data are omitted. NaSH does not force users to open up particular forms or follow a set path through the clinical interaction, as it is used in many diverse settings.

### Data quality

As well as completeness of any dataset it is important to review the quality of the data collected, including:

- **Error checking:** Data collection in NaSH is proforma-based and highly structured. NaSH has a high level of error checking on the majority of data fields of relevance to researchers. Some additional value lists may only become accessible to the user once a requisite preliminary data field has been completed and all data fields in NaSH allows the recording of 'not answered' and negatives. This aides in the interpretation of data, avoiding the difficulties that can occur when a system only allows affirmative answers to be recorded, which can make it difficult in retrospect to know whether the question was asked – a problem which was described by one of the clinics we contacted.
- **Referential integrity:** We found problems with referential integrity between forms in some key aspects of NaSH: one example is across recent and lifetime sexual history where recent sexual contact could be recorded with a partner of the same sex, at the same time as the lifetime sexual history indicates contact only with an opposite sex partner. There is no check back to highlight inconsistent data in real time to the user.
- **Computer assisted interviews (CASI):** There have been no studies in Scotland of NaSH comparing data quality and consistency from face-to-face history taking, paper recording and computerised record use. A number of clinics we contacted reported use of CASI by patients, allowing less time to be spent taking sexual histories and recording symptoms. During consultations, the clinician confirms the CASI data with the patient and enters new data on patient medical history, test requests, medication, treatment and diagnoses. CASI offers potential for effective electronic data entry and the use of such data in later research. So far no Scottish clinics have successfully deployed a CASI linked to NaSH.
- **Data audits:** Regular audits of data quality were emphasised by the other clinics we contacted as routes to improving and evaluating the system. Discussion, feedback, user surveys and regular meetings, and rigorous pre- and post-implementation testing were all mentioned as useful by the clinics we contacted.

### Accessing and using NaSH data

NaSH stores live data in a proprietary data format that is not accessible to end-users. For reporting purposes a large number of clinical data items are re-written in near-real time to an SQL reporting database in a format structured to make reporting easier. Reports for clinical use can be

compiled on all of these data items using the application's proprietary reporting tool, AdHoc. We identified four key issues to address in the use of NaSH for research:

- **Anonymisation:** For business reporting, anonymisation is achieved by interrogating a pre-specified virtual 'view', hiding all free text items and identifying data. The anonymisation algorithms were agreed by all Health Boards Caldicott Guardians (senior medical officers in the NHS who have responsibility to guard personally identifiable information). There are currently three geographic regional anonymous views covering West, East and North Scotland, and one national view. Additional anonymous views can be created by the application vendor and would be a possible way of creating a specific 'research' view with different anonymisation criteria to those needed for business level reporting. Any research project to access NaSH data would have to negotiate and pay for any changes needed to the data views, firewalls and additional storage space needed, particularly if this involved re-processing and re-coding of large data sets. Decisions would be needed about where to store re-processed data, given all NaSH data are currently held in the main NHS data centre under managed contract with Atos Origin Alliance.
- **Data over-writing:** We found a critical issue to do with over-writing of single-record form data in the reporting database, where only the most recently saved version of the data are reportable. This affects lifetime sexual history and lifetime social history such as alcohol and smoking. For example, a teenager with unsafe drinking recorded in 2011 may become an adult with safe drinking in 2013. Thus data are useful for cross-sectional analyses but less so for longitudinal studies. There is a need to develop a concept of archiving or storing key data items annually.
- **'Sense checking':** Those analysing data need to keep in mind the original purpose and setting of data collection. We noted that NHS Board analysts preparing service reports would be able to 'sense-check' the output against what was expected based on their knowledge of the service. This would be harder for researchers removed from the clinical setting.
- **Data retrieval:** This was a common problem among the review clinics we contacted. Data extraction and analysis is time consuming, complex and requires a high degree of technical proficiency and expertise handling large relational clinical datasets. The problems that ISD Scotland has had generating even basic data completeness reports so far illustrate this, but it is clear Scotland is not alone.

### Research governance

The need to comply with national regulations for electronic records systems, address the security and confidentiality of data, and gain informed consent were noted as issues by some of the clinics we contacted, but most clinics did not report experiencing any problems. Confidentiality is central to any use of sexual health data for wider research and there is a clear legal context around data sharing and confidentiality in the UK, with particular arrangements specific to Scotland that would affect using data from NaSH.

- **Ethical review:** The highly sensitive and personal nature of NaSH data suggests any research will require full review by the relevant Caldicott Guardian and Privacy Advisory

Committee (PAC), following the high impact research pathway suggested by the Scottish Informatics Programme (SHIP).

- **Anonymisation:** The existing anonymous data views were intended for use by NHS Board statistical analysts to support business planning and performance management and wider use of NaSH data for secondary research will require formal review of the anonymisation algorithms, the degree to which additional data may be required, the possibilities for, and the implications of (including risk of deductive) disclosure, and the use of data safe havens and one-way linkage to enable access.
- **Informed Consent:** NaSH does not contain any data items relating to consent to use personal data for research, secondary use or named patient use, or permission to contact patients by the clinics for research purposes. However, this could be incorporated into the NaSH consent model, alongside the other permissions already taken from patients when first registering. This would allow consenting patients meeting certain criteria to be contacted by researchers, e.g., for participation in sexual and social health interventions.
- **Longer term data retention:** Issues of longer term retention of NaSH data and availability to researchers out with the NHS have yet to be addressed, particularly if data are to be used for the purposes of longitudinal or retrospective studies. There is a tension between current NHS data retention policies and the value of preserving data for research purposes. There has been little guidance on whether electronically recorded data will be culled in the same way as historic paper records, and March 2015 (seven years after NaSH go live) will be a possible break point for this to be considered for NaSH.

## Conclusions and Recommendations

Greater use and interrogation of NaSH data could aid in the assessment of the indicators set out in the Scottish Government's Sexual Health and Blood Borne Virus Framework and have a key role in the future sexual health research agenda in Scotland.

It is clear that the quality of base data is a fundamental consideration in using routine data for research and, to facilitate data retrieval, it is imperative to identify the key information that should be on the system in numeric or coded form, as well as the user and technical proficiency required to maintain and access the data. There is potentially some conflict between the need for comprehensive and complete data for research purposes and the need for a routine clinical system to function in a routine way, within an acceptable timeframe and in a manner acceptable to patients and clinicians. Concerns over data collection, storage and retention should be considered within the context of the wider public health and research benefits of keeping a de-identified anonymised data set around key social and sexual risk variables.

To enable, and improve, use of NaSH data for sexual health research, we recommend the following:

- There is clear merit in making further use of NaSH to explore sexual health in Scotland, for population-level monitoring of sexual risks, and to document trends in recorded risk behaviour over time. As such, continuation of NaSH should be supported. Future system developments should be designed to make maximum use of the data to inform social and sexual health research as well as public health and epidemiology.

- Only some clinics currently include a question to gain consent to contact patients for research purposes in NaSH. Given the strengths of NaSH in identifying subgroups of interest we recommend this question be included in the routine demographic set as a searchable data item, as well as incorporated in all registration forms and on-line registration if and when this happens.
- Implementation of NaSH has focused on clinical utility and improvements in clinic process and no formal process has been established to review data completeness or quality at the national level. Data completeness and integrity checks should be developed, mapping this to the agreed minimum data set. The existing Lead Clinicians for each health board should be held accountable for oversight of these.
- Greater integration of kiosk-type or web-based forms should be considered; potentially allowing computer-based self completion. This would very likely improve data quality as the referential integrity checks that are missing in NaSH could be built into a CASI process.
- Data analysis in NaSH is complicated by the relational database and sheer scale of data items that may be drawn upon. A number of NaSH data items of social science research interest are stored in a 'single-record' form and can be amended at follow-up visits and only the latest entry is written to the reporting database. We recommend extracting these key data items into a protected data warehouse linking NaSH number with the values as known at a specific date, such as 31st December each year.
- The existing anonymous data views were intended for use by NHS Board statistical analysts to support business planning and performance management. Wider use of NaSH data for secondary research will require formal review of the anonymisation algorithms and the factors associated with this.

## I. Introduction

There is well established interest in the secondary use of health data recorded on electronic record systems for research purposes, and the UK is well placed to become a world leader in this approach (<http://www.mrc.ac.uk/Ourresearch/ResearchInitiatives/E-HealthInformaticsResearch/index.htm>). This is a fast growing field, subject to much development and review, with patient data recognised to be a valuable resource for research (Caldicott et al, 2013; Information Commissioner's Office, 2012; Thomas & Walport, 2008). In Scotland, the Scottish Informatics Programme (SHIP) has been funded by the Wellcome Trust, the Medical Research Council and the Economic and Social Research Council to provide a platform for the collation, management, dissemination and analysis of electronic patient records (EPRs). Among other things, SHIP aims to create a "research portal" for EPRs held by NHS Scotland and improve infrastructure for data linkage and sharing (see <http://www.scot-ship.ac.uk/c1.html>), and has conducted significant review of the research governance surrounding access to electronically recorded clinical data sets (Laurie and Syet, 2012).

Specialist sexual health services present specific challenges to electronic record systems implementation. These include:

- enhanced requirements from patients for discretion and sometimes anonymity, meaning standard hospital record systems and identifiers may not be appropriate;
- the need to register patients rapidly because many present on a walk-in basis;
- the need for rapid turnaround of laboratory results with easy access for patients because many patients desire to know results of their sexual health tests as soon as possible;
- the breadth of clinical care in sexual health, which includes detailed sexual history taking, minor operative procedures, under-16 child protection recording, prescribing and immediate administration of medication; and
- dispersed, often community-based locations that are neither acute nor primary care, and are sometimes delivered by third-sector (charitable) or voluntary groups.

In spite of these challenges various specialist sexual health services around the world have developed bespoke IT management systems, which make use of full EPRs, providing a booking and clinic management system and allowing recording of data on sexual risk factors such as recent and lifetime sexual history, lab test results, prescribing, and demographics. In Scotland, *Respect and Responsibility: A Strategy and Action Plan for Improving Sexual Health* identified the need for a standardised data collection system to support and monitor sexual and reproductive health services (Scottish Executive 2005a). As a result, the National Sexual Health System (NaSH), an electronic clinical record system, was developed and is now in use across all specialist sexual health settings in Scotland. NaSH is a specific development of Excelicare (AxSys Technology, UK). It is a full EPR, a booking and clinic management system, and links to key repositories such as the Community Health Index (CHI) database and the Scottish Care Information (SCI) Store. As of June 2013 it is live in all 11 mainland Health Boards across Scotland, although the extent to which the system is utilised varies from board to board (the three island boards do not have consultant-led specialist sexual health services and have not adopted NaSH).

While any IT solution must first provide the clinician and patients with the functions they need, sexual health EPR systems might also provide a wealth of sociologically important data.

Interrogating NaSH would enable researchers to make better use of existing sexual health data in Scotland, be cheaper than initiating large-scale surveys, give access to high-risk hard-to-reach populations, and provide the means to respond to immediate research questions. However, uptake of the electronic record varies across Scotland, and the data are recorded as part of a routine face-to-face consultation for clinical purposes rather than as part of an academic research study. Accessing and linkage of NaSH-derived data will also raise ethical issues due to the highly sensitive nature of the data recorded.

This review aims to assess the issues involved in accessing and using NaSH data for sexual health research in Scotland. The key research questions are:

1. What data are collected by NaSH, how comprehensive is the system, and what measures of data quality are available?
2. What would be the key issues to address in routine use of NaSH data to better yield benefits for sexual health research?

Out of scope of this review is any assessment of the clinical benefits or otherwise of implementing NaSH. The Clinical Portfolio Management Group established a Short Life Working Group in November 2011 to review the benefits, technical challenges and financial implications of NaSH. As of June 2013, the findings are yet to be publically released, but this Report will complement these.

## 2. Methods

The scoping review was conducted in three stages: a policy review of NaSH documentation; a review of EPR issues reported by an international selection of clinics known to be using computerised clinical systems; and a review of more general methodological issues related to the use of EPR.

### 2.1 Policy and development review

Formal policy documents relating to the development of NaSH, the implementation of EPRs and other relevant developments and issues (such as information governance) in both sexual health settings and the NHS in Scotland in general were obtained from the Scottish Government website.<sup>1</sup>

In November 2011, we gathered material on the planning, development and early implementation of NaSH from the dedicated NaSH website,<sup>2</sup> including the original Project Brief, the Statement of Requirements, minutes of meetings of the Project Board and the Reference/User Management Group, documentation on benefits and project assurance, and general background information about the project. Additional documentation was provided by collaborating individuals and organizations: Dr Andy Winter (AW), of NHS Greater Glasgow and Clyde and chair of the NaSH User Management Group; the Information Services Division Scotland (ISD), the health service data collection organisation in Scotland; and Dr Steve Baguley of NHS Grampian. Additional background information was accessed via the websites of a number of NHS Scotland organisations and divisions: ISD<sup>3</sup>, HPS<sup>4</sup>, Healthcare Improvement Scotland (HIS)<sup>5</sup>, NHS National Services Scotland (NHS NSS)<sup>6</sup>, the eHealth Programme<sup>7</sup>, and Information Governance<sup>8</sup>. A total of 124 documents were reviewed (the full list is available from the first author), and the following data were extracted:

1. NaSH background (context / why the system is needed / objectives)
2. NaSH clinical functions (what NaSH could be / is being used for in clinical practice)
3. Other functions of the system
4. Data collection (what data could be / are being collected)
5. Data quality and validity measures / auditing processes

---

<sup>1</sup> <http://home.scotland.gov.uk/home>

<sup>2</sup> <http://www.nash.scot.nhs.uk/>. All documents from the NaSH website were downloaded in November 2011. As of September 2012 the website is no longer accessible.

<sup>3</sup> <http://www.isdscotland.org/>

<sup>4</sup> <http://www.hps.scot.nhs.uk/>

<sup>5</sup> <http://www.healthcareimprovementscotland.org/home.aspx>

<sup>6</sup> <http://www.nhsnss.org/>

<sup>7</sup> <http://www.ehealth.scot.nhs.uk/>

<sup>8</sup> <http://www.knowledge.scot.nhs.uk/ig.aspx>

6. What data cannot be / are not being collected and why
7. Practical considerations / issues affecting staff using the system
8. Systemic or external security/confidentiality arrangements
9. Wider ethical issues

## 2.2 Review of use of EPR systems in other sexual health clinics

We identified 10 clinics known by one author (AW) to have a particular publication record and/or interest in computerised clinical data systems (Table 2.1). Five were in the UK, two in Australia, two in the United States and one in south-east Asia. It is likely that there are clinics in other locations using electronic records or computerised systems, but these were the ones we identified initially as being of particular interest. Eight of the ten clinics reported having a computerised sexual health recording system in place. Two clinics noted that they did not have a current EPR/computerised system in place: the Courtyard Clinic and the public STD Clinic in Seattle. The Courtyard Clinic used paper records, alongside an electronic booking system, but planned to implement electronic recording in May 2012. Seattle described a process whereby paper charts were used and data from these were later entered into an electronic database, which could then be used for clinical and research purposes, with data going back to 1993. They also planned to move to electronic recording in the future, potentially in 2012. Seattle did use a Computer Assisted Self-Interviewing (CASI) system to gather demographic, clinical and sexual risk information from patients before seeing a clinician.

**Table 2.1: Sexual health clinics included in the Review**

Clinic name	City	Country
Courtyard Clinic	London	UK
Mortimer Market Centre	London	UK
Patrick Clements GUM Centre	London	UK
West London Centre for Sexual Health	London	UK
Whittall Street Clinic	Birmingham	UK
Melbourne Sexual Health Centre	Melbourne	Australia
Sydney Sexual Health Centre	Sydney	Australia
Baltimore City Health Department Public STD Clinics (Druid Clinic and Eastern Clinic)	Baltimore	USA
Public Health – Seattle and King County STD Control Programme (Public STD clinic at Harborview Medical Centre)	Seattle	USA
Department of STI Control (DSC) Clinic	Singapore	Singapore

Information about the use of computerised sexual health systems were requested from the clinics noted above to examine key issues around the data collected, coverage and quality, practicalities and ethical issues. Emails were sent to a named clinician for each clinic, requesting the following:

- When was a computerised sexual health system introduced in your clinical setting?
- What data are collected by your system?
- How are data collected (i.e. patient self completion/ data entry in real time by clinician)?
- What is the coverage and quality of the data collected, how is this monitored and how have you made improvements?
- What, if any, are the practical issues you have found to be associated with the use of data from the computerised system for sexual health research?
- How have you overcome any problems experienced in the use of your computerised data for research purposes?
- What, if any, ethical issues have you faced in relation to the use of data from the computerised system for sexual health research?
- Are the data used to inform service delivery and/or public health planning?
- What, in your view, are the key advantages/disadvantages of using a computerised sexual health system?

Responses were received from all clinics and the replies were entered into an Excel spreadsheet for comparison. Our review is limited to the information provided by each clinic, but omission should not be taken to equal absence, and in the Results we concentrate on identifying the dominant themes that emerge from cross-clinic comparison. A table showing the year of system implementation, data collection methods, data collected, quality assurance and monitoring across clinics is included in Appendix 2.

### **2.3 Review of general methodological issues**

We also examined general methodological issues around the use of routine data for research, including research in other health areas. 34 papers were identified from initial literature searches of the Pubmed, Embase and Web of Knowledge databases using the following MESH terms:

- Data Collection
- Data Collection/utilization
- Data Interpretation, Statistical

- Health Care Surveys/methods
- Medical Records Systems
- Medical Records Systems, Computerised
- Sexual Behaviour/statistics & numerical data
- Sexually Transmitted Diseases
- Venereology

In addition, the following free text phrases were used: *routinely collected data*, *routinely collected*, *coding*, and *data collection*. Of the 34 papers returned, 16 were excluded for reasons including: being too topic-specific (e.g., quality of care) or having too narrow a focus (e.g., data coding only); no discussion of methodologies; and/or not involving routine data. Of the remaining 18 papers, one further paper was excluded because it dated from 1997 and discussed methods of data extraction which could now be considered outdated. A further paper was identified in the course of the extraction and was subsequently included. Five additional articles, editorials and letters were provided to us by Dr Jackie Cassell of Brighton and Sussex Medical School, giving an overall total of 23 papers. Included papers were reviewed for issues on: data collection; data management (including storage, IT, practicalities and quality assurance); data extraction and analysis; data governance (including consent, ethics, confidentiality, access to data, and data sharing); and the opportunities and limitations that were commonly linked to research uses.

## 3. Results and Discussion

### 3.1 What data are collected by NaSH, how comprehensive is the system, and what measures of data quality are available?

In the following section, we consider what data are collected in NaSH, how comprehensive is the system, and what measures of data quality are available. Examples and comparisons of how other clinical systems have addressed issues and problems, and from the review of general methodological issues, are presented throughout.

### 3.2 Data collection in NaSH

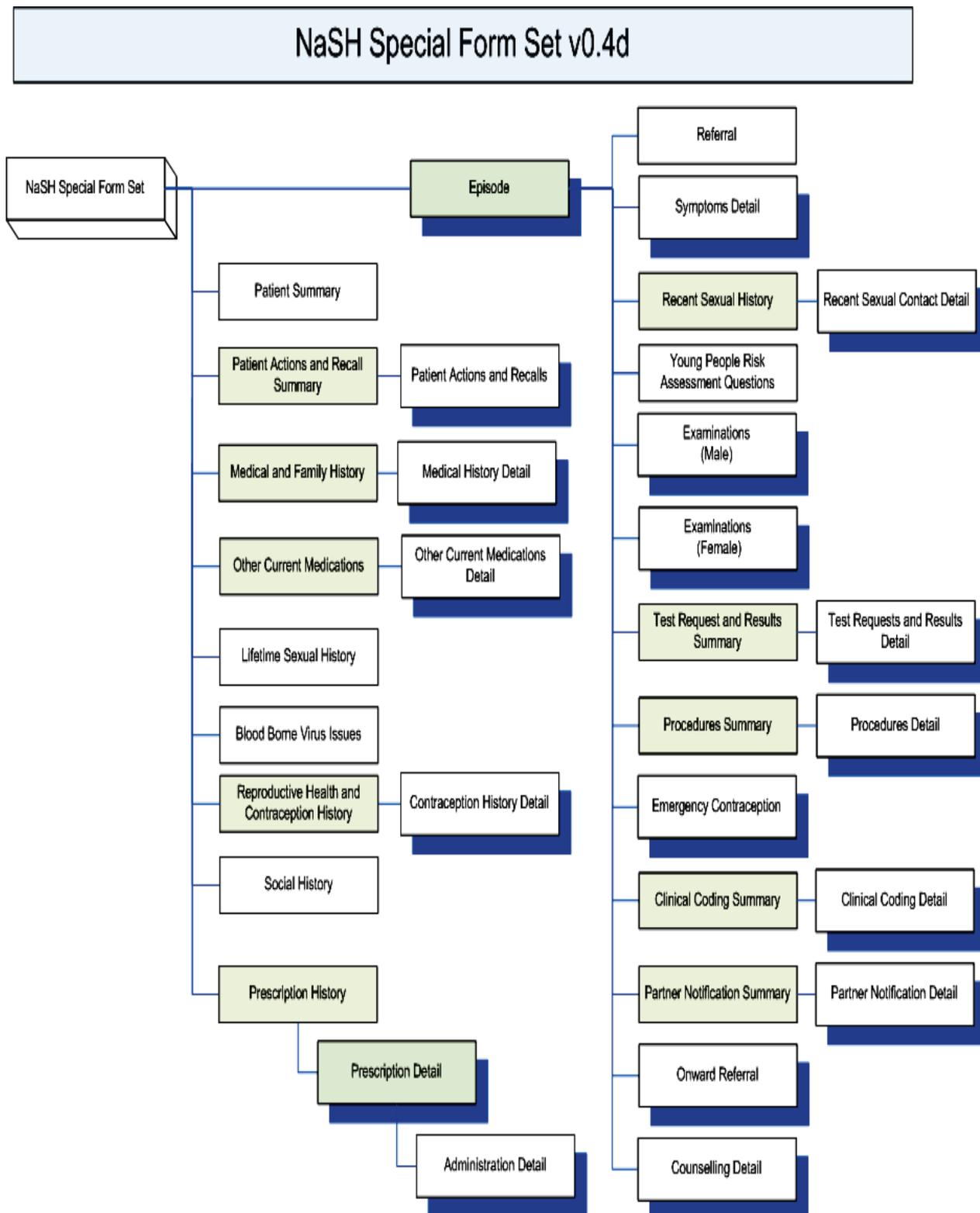
Data collection in NaSH is intended to be done in real-time as the clinical or administrative user navigates through the system. The system supports the collection and recording of data in relation to the following processes:

- Creating and setting up clinics
- Appointment booking
- Patient registration
- Medical history
- Lifetime sexual history
- Blood borne virus risk history
- Reproductive and contraception history
- Social History
- Episode data, including recent sexual risk data, partner notification, examination findings
- Laboratory tests
- Procedures
- Prescriptions and drug administration
- Clinical coding (this was the old process for STISS coding within NaSH)
- Onward referrals
- Actions and Recall
- Counselling
- Free text clinical notes

Special forms have a range of clinically relevant data items on each page. The user is free to complete ones of interest and to record data in several areas without being tied to a single route through the system. This approach has encouraged flexibility, but it also has inherent risks for incomplete data collection. The core of the database in terms of clinical, behavioural and lifestyle data is the NaSH Special Form Set, which covers an extensive range of variables including medical and family history, medications, lifetime and recent sexual history, blood borne virus (BBV) issues, reproductive health and contraception, social and lifestyle factors including smoking and alcohol intake, test requests and results, patient actions and recalls, prescriptions, symptoms, physical examination details, partner notification, counselling, and referrals. The information entered in each special form in the set is linked under an overall Episode form. This is the master form for each episode of patient care. It acts as a central point for users and provides access to the other special forms within the NaSH system. The Special Form Set is represented in Figure 3.3. The

forms include free-text fields in which additional details and notes can be typed as necessary, and number fields where digits can be entered as required (as for the number of sexual partners). Otherwise the bulk of clinical data entry is through the selection of pre-set coded values from drop-down lists. There are very few mandatory data items specified, in order not to lock users.

Figure 3.3 NaSH Special Form Set<sup>9</sup>



<sup>9</sup> Figure taken from Winter (2010: 50).

The special form set is a combination of multi-record and single-record forms. The Episode form is multi-record. Users are able to create records of subsequent episodes of care for the same patient by completing another Episode form. Single-record forms (such as medical or social history) only have one record completed per patient, although the information entered into the form can still be edited at a later time. Only the *latest* visible data are written to the reporting database, with previous data recorded being visible to users by calling up previous iterations of the form on the screen. This is incredibly important because it means only the most recent data are recorded in a retrievable way, and any change over time cannot be examined; hence, making longitudinal analysis impossible.

### **3.2.1 Recording sexual history**

Sexual history recording involves three core forms: the Lifetime Sexual History form, the Recent Sexual History form and the Recent Sexual Contact Detail form. The Lifetime Sexual History form is single-record and records the last recorded answers to key risks, such as whether the patient has ever had sex with another person of the same gender. This can be updated at any time, but only the last entered data are copied to the reporting database. The Recent Sexual History form should be completed for each episode of care, and records summary information on recent sexual partners, such as date of last sex and number of recent partners. Further detail for each partner can be entered in Recent Sexual Contact Detail forms (more than one of which can be created for each patient per episode of care). Data items include the gender, age and place of residence of the patient's most recent sexual partner, the type and location of sexual activity and protection used.

International comparison could also be facilitated by the collection of comparable behavioural data and the use of comparable time scales. A study by Fairley et al. in 2010 reviewed the development of a core sexual history in sexual health clinics in the United States, the UK, Australia and New Zealand. There were key areas of convergence in the questions involved in a sexual health history in all four countries: gender of sexual partners, number of partners (albeit across varying time periods), types of sex (oral, anal, vaginal), condom use, pregnancy intent and methods of contraception. Fairley et al. also pointed out that the time periods covered in sexual histories can vary. This is borne out in the responses we received from individual clinics. The Public STD Clinic in Seattle, the Melbourne Sexual Health Centre and Sydney all stated that they collected the number of partners over the previous 12 months. While the long-range time period was constant, the short-range period varied. Seattle collect partner numbers for the previous two months, but Melbourne and Sydney use the slightly longer period of three months.

### **3.2.2 Recording STI test results**

It is possible to import laboratory test results directly into NaSH via a link with each Board's SCI Store, an electronic data repository developed by NHS Scotland which allows patient information to be shared between systems in a Board. There is a separate instance of SCI Store in each board in Scotland. It is used to share patient demographics, lab investigation reports, radiology reports, treatment logs, clinical documents, and information on patient admissions, discharge and transfer<sup>10</sup>. SCI Store uses the CHI number as the primary identifier, but in most Boards anonymously identified sexual health specimens with NaSH 'AN' numbers are also permitted. As at June 2013, five Boards had integrated SCIstore and NaSH, with a further three to go live in 2013. The main challenge to laboratory results integration is that SCI Store only holds the data in the way they are

---

<sup>10</sup> See [http://www.sci.scot.nhs.uk/products/store/store\\_main.htm](http://www.sci.scot.nhs.uk/products/store/store_main.htm) and <http://www.sci.scot.nhs.uk/products/store/General/SCI%20Store%20-%20Product%20Description.pdf>

given, and many laboratory systems output lab results for microbiology in a non-structured format. However, results can be entered by hand onto the Test special form, so a complete data set of sexual infection test results is available by interrogating the NaSH form sets in most Boards.

### **3.2.3 Recording social risk factors**

NaSH has a specific form set to encourage recording of social risk factors, including smoking history, alcohol history, recreational drug use, gender-based violence, and accommodation concerns. This is on a single-record form so only the latest entered data are written to the reporting database with previous entries visible to clinicians if required. There is also an under-16 year old specific reporting form which records child protection concerns and parental involvement.

SCI Gateway<sup>11</sup> is the primary method of primary care written referral into secondary care. This does not currently integrate into NaSH, although the electronic referrals are usually copied and saved into NaSH as a non-searchable PDF document. Thus, information in a Gateway referral about previous health and social risk (for example smoking) cannot be electronically mapped into the relevant NaSH data fields. Furthermore, many patients are simply verbally advised to attend sexual health, so it is unwise to rely on GP derived data for risk factors solely by examining the written referrals.

### **3.3 Anonymised data view**

For reporting purposes a large number of clinical data items including free text notes are transformed from the proprietary live database structure and written in near-real time to the reporting database with a more intuitive table structure, which is available in a range of views. Board staff working in sexual health can access the entire data set using a vendor-supplied database viewer ('AdHoc') and a range of dynamic live 'PRISM' reports are used to drive workflow on the system. Business reporting makes use of anonymised data views aggregated at regional level and the earliest of these was developed for the West of Scotland Managed Clinical Network. Two further views have been created for the East of Scotland and the National ('ISD') view, while the North of Scotland view is pending implementation. Anonymisation is achieved by removing all free text items and any data that might identify an individual. For example, server-side scripts map current postcodes to lookup tables reporting geographic residence down to the level of community health partnership, and exact SIMD quintile can also be returned. Scripts also generate 'age at' settings for key items such as prescriptions and laboratory tests and suppress date of birth.

The key focus for the development of the anonymous views was for business reporting to enable service planning especially at a regional level. It is important to understand that the view is a dynamic representation of real-time, changing data, not a stored data set. Subject to appropriate ethical review, this data view could also be a useful resource for sexual health research. When considering the use of NaSH data for research the following issues should be addressed:

- The level of detail to which geographic location is reported, e.g., more utility might be gained mapping down to intermediate zone but this increases risk of deductive disclosure
- The choices made about which data items are reportable on the anonymous view

---

<sup>11</sup> See [http://www.sci.scot.nhs.uk/products/gateway/gate\\_desc.htm](http://www.sci.scot.nhs.uk/products/gateway/gate_desc.htm)

- Further verification of the anonymisation algorithms, which were approved for access by health-board and ISD employed business data analysts but may require additional stringency for release of data sets to approved researchers.
- Consideration of archiving and storage of complete data sets. So far there have been no proposals to archive or copy off data for key areas, but this might be considered by ISD to facilitate calendar year reports. Further consideration of how to ‘warehouse’ data for reporting purposes and possibly for future research is warranted. As noted by one of the clinics we contacted, despite having a ‘huge amount’ of epidemiological and behavioural data they wished to make use of, they were simply finding it hard to get the data out.
- For researchers to use the data, an understanding of how the original data were collected and under what conditions and assumptions is required. Business analysts working for the service will ‘sense-check’ their reports with lead clinicians and managers, and this reduces the chance of misunderstandings about data items purpose and interpretation. An example would be that certain questions on special forms are only revealed to clinicians based on a positive response to a previous data item, but this is not explicit in the existing data model supplied. Therefore, it might appear that questions are not being completed when in reality they were not applicable.

### 3.4 Data Linkage

NaSH allows patient-centred choice of whether to use an anonymous identifier (a unique NaSH number generated within the system) for their lab tests and records within the system or to allow the use of their unique NHS identifier, or CHI number, which could facilitate record linkage.

A model was developed with three options for patient consent to the use of his/her CHI number: full identification with use of CHI number; identification for CHI demographics with anonymous testing; and assumed identity with anonymous testing (the anonymous NaSH number is then used for internal and external communications) (Paterson 2008b). Outside of NaSH, there is no way of linking a patient’s CHI number to their NaSH number. In addition, if a patient initially agrees to the use of their CHI number they are still able to withdraw their consent later on. In such cases, the NaSH number is used for any subsequent communications. This patient-centred model has proved successful in at least some boards, with around 76% of patients agreeing to download of CHI demographics at registration in NHS GG&C in 2011<sup>12</sup>. Together with the postcode finder this means that in comparison to historic stand-alone sexual health systems, basic demographic recording is very good, with a high proportion of patients in at least Greater Glasgow and Clyde (94%) having valid postcodes<sup>13</sup>. This provides the possibility of linking a significant proportion of sexual and social health items held in NaSH, at least for the subset of patients with CHI identifiers stored on the system, subject to approval and consideration of the one-way encryption and linkage process.

Currently, for obvious reasons, the CHI number is not included in the anonymous data view. Pseudonymisation in NaSH is currently achieved by using the patients NaSH number. However,

---

<sup>12</sup> As shown in NHS GG&C Clinical Governance Snap Shot for Hub, 01.01.11 - 31.12.11 (Personal communication from AW).

<sup>13</sup> NHS GG&C Clinical Governance Snapshot for Hub pan healthboard, 01.01.11 to 31.12.11.

there have been significant advances in approaches to anonymisation since the NaSH data views were constructed and approved. These include the 2013 Caldicott Review on Information Governance, which recommended that personal, confidential data or data from which individuals could potentially be identified should be only made available via secure data 'safe havens' (Caldicott et al, 2013). The current NaSH anonymous data views still contain individual's single NaSH identifier and as such should be regarded as data 'for limited disclosure', in that individuals could be 're-identified or de-anonymised' (Caldicott et al, 2013; Information Commissioner's Office, 2012; Thomas & Walport, 2008).

### **3.5 ISD minimum dataset**

In 2011, ISD proposed implementation of a minimum input criteria for NaSH to contribute to data for the national sexual health information reports produced by ISD. The minimum input criteria were designed to maintain ISD's existing level of sexual health reporting – i.e. to continue to capture what had been captured in STISS – but to expand the application of this to integrated settings and to support reporting on the national key clinical indicators for sexual health (Table 3.1). There is no requirement to complete every field for each patient, as not all fields will be relevant in all cases (functional non-completion), but every applicable field should be recorded every time.

**Table 3.1 ISD minimum input criteria<sup>14</sup>**

Section/screen	Variable/Field	Notes
Patient Details	Date of Birth	ISD receive these data as a derived Age field.
	Sex	
	Postcode	ISD receive this as derived Deprivation Category and Community Health Partnership/NHS Board fields.
	Ethnic Category	
	Disability/ Health Conditions	Not required to include chronic health conditions if they cause minimal disability (e.g. asthma).
Episode details	Start Date	
	Clinic Code	
	Main Reason for Attending	
	Secondary Reason for Attending (if required)	
Lifetime Sexual History	Gender Previous Sexual Partners	ISD receive these data as derived Lifetime Sexuality Status.
Blood Borne Virus Issues	Injecting Drug Status	
	HIV Test Status	Completion required only for patients known to be HIV positive. Recording this will allow the number of patients with known HIV attending for STI screening to be measured.
	Vaccination History: Hepatitis B Vaccination	Completion only required for IDU and MSM patients. This allows calculation of proportion of MSM without prescription of Hep B vaccine who remain unvaccinated.
Recent Sexual History	New Partner in last 3 months/ last 12 months	Allows calculation of STI rates and testing in those at recent risk.
	Gender of Sexual Partner	Allows acute STI to be ascribed to same-sex or different-sex partnerships.
Test Request and Result Details	Test	The Patient Order function should be completed for the following tests: Chlamydia, gonorrhoea, HIV, syphilis, HSV, Hep B and Hep C.
	Date of Test Request	
	Interpreted Result	All positive results should be entered.
Prescription Detail	Drug Name	As a minimum, this should be recorded for treatment of herpes and warts and for provision of contraception, particularly LARC.
	Start Date (date of prescription)	
Patient Procedures: Implant	Date of Procedure	Details of both fittings and removals of contraceptive implants or devices should be recorded as well as details of prescriptions for these. This will enhance reporting on the uptake of contraceptive methods.
	Type of Procedure	
	Reason for Removal	
	Time in Place	
Patient Procedures: IUD/IUS	Date of Procedure	See above.
	Type of Procedure	
Patient Procedures: Vasectomy	Date of Procedure	

<sup>14</sup> Derived from Information Services Division 2011c.

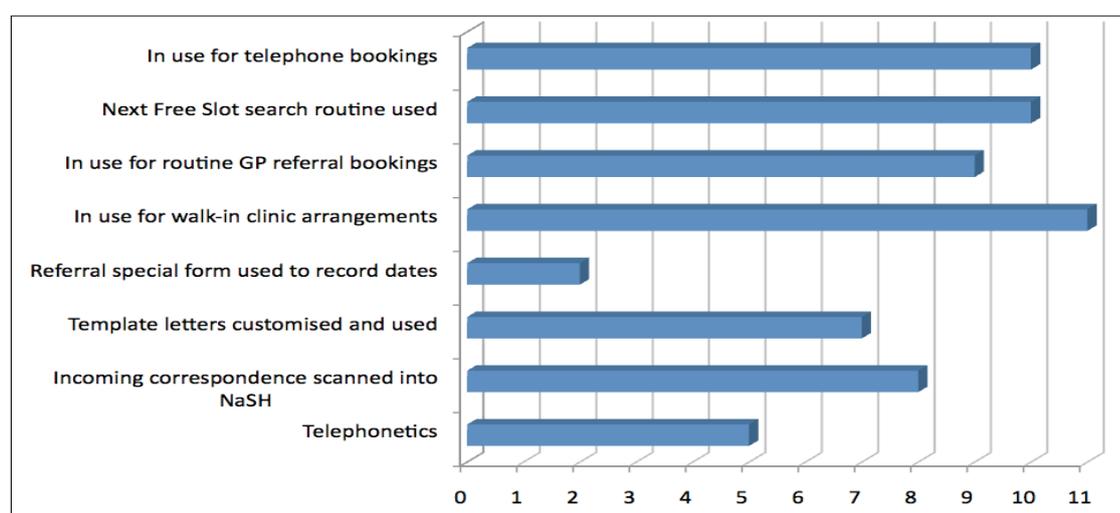
NaSH was built to record a high level of detail on social and lifestyle risk factors, but the majority of these (except for sex, age, postcode and ethnicity) are not currently included in the ISD minimum dataset. This limits the possibility of conducting more sophisticated and inclusive analyses of clinical and sociological variables. Social and lifestyle risk factors impact on sexual health and can be extremely valuable when linked with epidemiological and behavioural data. The range of data that other clinics using EPR reported collecting over and above core demographics (name, gender, date of birth) includes the following: the patient's relationship status, domestic violence screening, smoking and alcohol use, ethnicity and occupation. Social factors that could be collected in NaSH include alcohol use, smoking, substance misuse, accommodation status, violence and abuse and eating disorders. Analysis of these factors at a national level could be difficult and would have to take into consideration a number of issues of data completion and completeness.

### 3.6 NaSH data completion and completeness

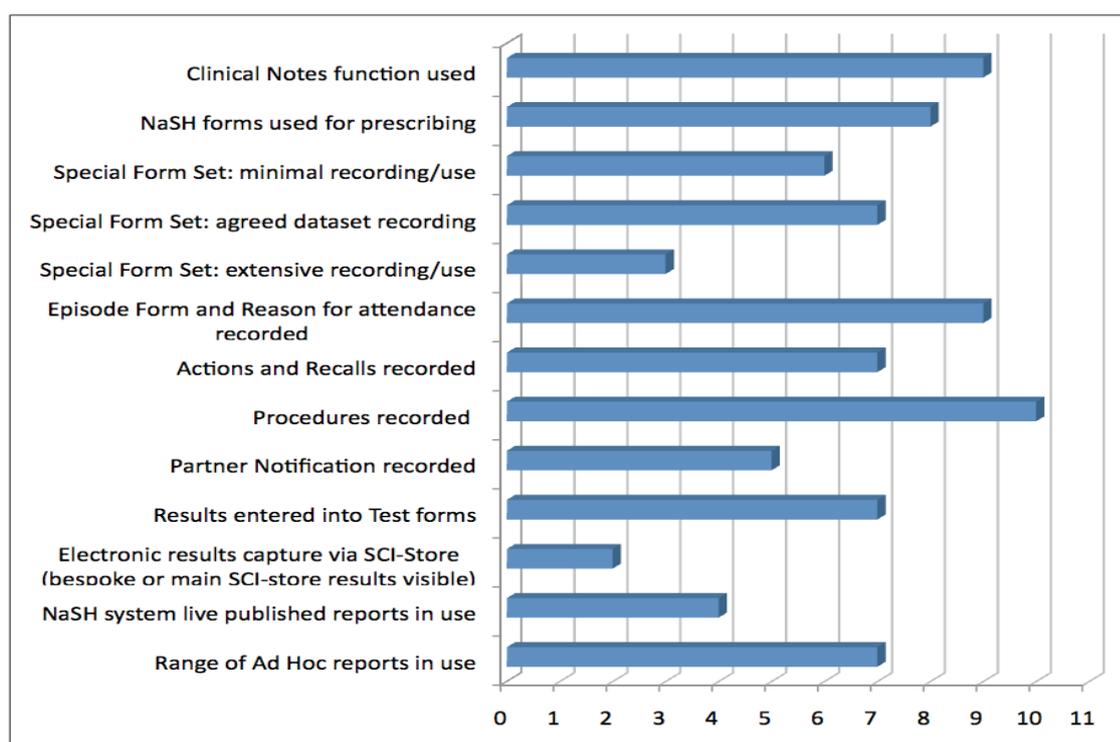
A national overview of data completeness is still lacking as at June 2013 due to difficulties in ISD accessing and analysing the national anonymised NaSH data view. This is essential to determine the utility or otherwise of using routinely collected NaSH data for sexual health research. In the meantime, we assessed data completion and completeness at Board level.

As of October 2011, NaSH had around 700,000 patients registered and 2,300 user accounts (Naughton et al. 2011). It records around 300,000 visits a year or more. Usage as at Figure 3.1 shows that the booking elements of NaSH are well used, while the referral and correspondence elements are less so (data from October 2011)<sup>15</sup>. Figure 3.2 shows that the majority of boards (7 or more) were completing special form datasets, recording prescriptions and procedures, and entering laboratory test results. There is less recording of partner notification and most boards were struggling with the electronic laboratory links.

**Figure 3.1 Scheduling, registration and administrative functionality across NHS Boards**



<sup>15</sup> Data derived from NaSH - Health Board Adoption at October 2011 (Personal communication from AW).

**Figure 3.2 Clinical, recording and reporting functionality across NHS Boards**

NHS Boards with NaSH in place were to begin completing the minimum dataset from January 2012 and returning data on the required fields to ISD. The minimum dataset has the potential to ensure a standard level of data recording cross-nationally as well as within individual health boards and means that it could allow the reliable analysis of data for research at the national level in the near future. However, use of NaSH in real time has been problematic in some locations due to connection speed, and the actual design of the system itself (Steve Baguley, personal communication). This is cited as the main reason why use as a full EPR has not been taken up in some areas. This is not an issue unique to NaSH. Other clinics reported specific technical issues included computers crashing before data could be saved, system crashes following upgrades, and having insufficient storage and network capacity. These were cited by clinics in the UK and Australia as well as in the United States. One clinic reported that there was a back-up paper record system in case of occasional computer malfunctions (with all data later re-entered into the computer).

To counter these problems, some board areas have adopted paper or third-party computer-based self-completion forms, whereby clinical staff transcribe positive information from the self-completion form which is later scanned to form a record. The problem with this approach in terms of data analysis is that lack of risk and negative findings are not recorded. Furthermore, all data may not always be completed routinely in practice in all health boards, meaning that comparison across health boards may be problematic. This variation is inevitable in a large national system, and to some extent reflects different priorities within the boards towards monitoring of different areas, for example for gender-based violence and alcohol risk reduction. This needs to be understood when analysing national data.

The two main data collection methods reported by other clinics using EPR were real-time electronic entry and the use of patient CASI combined with data entry by clinic staff (either

clinicians or others such as reception or laboratory staff). However, even in settings where the bulk of data is collected on paper first, the use of computerised patient self-completion represents an element of real-time electronic entry. Seattle noted that following their introduction of CASI in 2010 clinicians 'probably' spent less time taking sexual histories and recording symptoms, since these are now captured by the self-completion system. At the time of writing, Sydney used CASI only in their Xpress clinic for asymptomatic patients with no other concerns, but they planned to expand CASI in future to allow patient self-registration and entry of demographic data. This is similar to the use of CASI at Melbourne, where general patient information and associated risk data is self-completed by the patient at registration before seeing a clinician. During consultation, the clinician confirms the CASI data with the patient and enters new data on patient medical history, test requests, medication, treatment and diagnoses. A number of studies have shown CASI to be effective in clinical setting and to be acceptable to patients and clinicians, with no adverse effect on clinical output (Ghanem et al. 2005; Richens, et al. 2010; Tideman et al. 2006; Tideman et al. 2007; Vodstrcil et al. 2011). CASI and ACASI offer potential for effective electronic data entry and the use of such data in later research.

The NaSH system can generate local reports that show how well data items have been completed, making it possible for services to assess their own dataset compliance and identify areas of high and low completion. In NHS GG&C, completeness was generally good across their minimum dataset in 2011<sup>16</sup>. For example, social history forms were completed for over 80% of all existing male and female patients in 2011 and between February and September 2011 domestic violence/abuse status was recorded for over 78% of existing female patients and 84% of new female patients (the figures were slightly lower for men, with status recorded for 68% of existing and 78% of new male patients). This demonstrates that there are clear issues with differential completeness. In NHS GG&C, young people (those aged under 20) are more likely to have more complete recording of social risk and men are more likely to have their lifetime sexuality recorded. Attempts to use such data at a national level would require further research to assess whether they are part of a wider pattern, or are indicative of a wider tendency to record risk most completely for patients who are already perceived to be at risk: *targeted completion*, in other words.

The flip-side of targeted completion is what might be called *functional non-completion*. In NHS GG&C in 2011, it is notable that 31% of episodes (the largest single group) were recorded as having been prompted by routine contraception issues. In day-to-day clinical practice, it may be seen as unnecessary or impractical to go over recent sexual history with a return patient attending for a routine implant removal, for example; this is functional non-completion. Both targeted completeness and functional non-completion can be contrasted with completion for completion's sake, or *dysfunctional completion* – collecting and recording data when there is no actual, clinical need, and which may end up impacting adversely on the value of the data and later research. Clinicians may be disinclined to ask, and patients may be unhappy with answering, questions that do not appear to relate to the problem or issue for which they are attending the clinic, even if there is public or population health merit. The problem lies in the decision as to when functional non-completion is appropriate, for which data items, and with which patient groups.

---

<sup>16</sup> Clinical Governance Snapshot for Hub pan-healthboard, 31.01.11-21.12.11.

### 3.7 Measures of data quality

As has been discussed, clinical data collection in NaSH is proforma-based and highly structured. In many instances, it is possible to automatically 'pull-through' relevant data that have been entered at other points in the record. This saves manual re-entry and also militates against repetition error – although if the data have been entered incorrectly in the first place it will simply compound the problem by replicating the mistake until it is noticed.

NaSH has numerous features designed to prevent entry errors. Some additional value lists may only become accessible to the user once a requisite preliminary data field has been completed: for example, in the Recent Sexual History Contact Detail form, the list of specific barrier methods (e.g. condoms) will only become active if the previous data field indicates that a barrier method was used. In turn, the field relating to successful or unsuccessful use of any barrier method will only become active once a barrier method has been entered. In theory this makes it difficult for contradictory data to be recorded within forms – so there should be no risk of the record indicating that the patient did not use a barrier method at the same time as indicating that he or she successfully used condoms. Fields remaining inactive until the previous field has been completed can also prevent accidental non-entry of data, or items being missed, because it will not be possible to proceed through the form until this is rectified. The use of compulsory fields in the computer system which must be completed before proceeding was one measure used in clinics elsewhere, while pop-up reminders and alerts were another alternative.

NaSH also allows the recording of negatives. It is possible to record whether the answer to a particular question was No, whether the question was not asked, whether the patient did not disclose any answer or did not know the answer. This helps to prevent the loss of data that can occur when a system only allows affirmative answers to be recorded, and which can make it difficult in retrospect to know whether the question was asked – a problem which was described by one of the clinics we contacted. This is a result of systems having evolved with a tick box, or similar, to indicate a problem. An unticked or unchecked box could mean either the problem was absent (e.g. no symptoms) or the question was not asked or forgotten, but there would be no way of knowing which interpretation was accurate. As noted above, negative recording could vary across health boards and would have to be considered prior to using the data. One of the other clinics using EPR that we contacted described having made changes to many of their questions so that there was a 'No' as well as a 'Yes' answer option. This demonstrates the importance of making sure that questions are designed not just to ask the right things but to allow the fullest possible answer, and also that there may be a need to review and revise NaSH to ensure the best use of the data collected.

In reality, referential integrity can easily be undermined by human error in combination with certain weaknesses in the system. For example, there is no way of validating data recorded across recent and lifetime sexual history. This leaves scope for between-form 'inconsistency creep'. Recent sexual contact could be recorded with a partner of the same sex, at the same time as the lifetime sexual history indicates contact only with an opposite sex partner. This makes the existence of a strong auditing capability all the more important. Any editing or deletion of records in NaSH is logged in the audit trail built into the system and is attributed to the user. The audit trail also supports post-hoc quality assurance. Reports such as those discussed above produced by NHS GG&C also allow data gaps to be identified, which can then be traced back through the records themselves if necessary.

Regular audits of data quality were common across the other clinics we contacted, and these chiefly consisted of record or file reviews. At Sydney, all files of new staff were manager-reviewed until the error rate in data entry fields was below 5%. The clerical manager randomly audited data entry clerking accuracy. Random quality assurance was carried out on medical records and rotated among all staff. This also encompassed clinical approaches. One of the American clinics stated that medical directors audited a sample of around 10-15% of records to ensure data quality as well as quality of care. Previously, all records were audited, but this had recently been decreased due to provider consistency. They also noted that they had a full-time member of IT staff who ensured integrity of data entry and storage. In contrast, one of the UK clinics described how they wanted to do more in terms of assurance and monitoring but were restricted by a lack of IT support.

As routes to improving and evaluating the system, discussion, feedback, user surveys and regular meetings, and rigorous pre- and post-implementation testing were all mentioned by the clinics we contacted. Non-systemic methods for data quality assurance included cross-comparison between records: checking the accuracy of coding by comparing lab records, or checking free text entries through paired note reviews. This is essentially a form of data triangulation. Where CASI is used alongside face-to face consultation to collect patient information, this also constitutes methodological triangulation. Both can function to promote data quality.

### **3.8 What would be the key issues to address in routine use of NaSH to better yield benefits for sexual health research?**

Here we discuss the practical and ethical issues associated with the use of data from computerised systems for sexual health research, and examine how these might affect the potential for research using NaSH. We draw from the NaSH documentation, input from the participating clinics, and methodological papers and studies that focus on the use of routine data in research. Table 3.2 provides an overview of the key advantages and disadvantages of using data from a computerised system for sexual health research. These have also been noted elsewhere during implementation of EPR (Brooks et al. 2008).

### **3.9 Data collection and completeness**

From the perspective of secondary-use and research, it is worth reiterating how important it is that data are collected as completely, correctly and with as much standardisation as possible. In thinking about the use of NaSH data, this would translate to always being aware of the origin of data, how they have been collected and how this might have affected data quality.

**Table 3.2 – Using data from a computerized system for sexual health research**

<b>ADVANTAGES</b>	<b>DISADVANTAGES</b>
Instant/more rapid access to data	Data retrieval can be technically difficult; may be legal or ethical barriers to access
Saves physical storage space	Still costs associated with data storage
Saves time and resources	Purchase and maintenance of software/hardware Staff training
Data potentially of better quality, greater standardisation	Missing or incomplete data and unreliability of self-collection
Analysis is more timely, more sophisticated, easier	Potential for misinterpretation
Comprehensive range of variables covered	Cannot cover everything – data fatigue; changes in variable definitions over time affect how data can be used later
Can cover large populations	Results still may not be generalizable
Bringing in less traditional areas of sexual health research e.g. associations between partner violence	These areas may be affected by incomplete/differential recording
Allows longitudinal research over time	Large volumes of data may be daunting and difficult to process; subject to time limits for retention, just as paper records
Less potential for record mix-ups / lost notes	Data loss or corruption due to system problems
Increases scope for other uses of data	Danger of unfocused fishing expeditions
Enhanced security	Perception of vulnerability (e.g. to hacking)
Data linkage	

Simple human errors or omissions at collection stage, or variability in completeness across different settings, can result in missing or uneven data and affect everything that is done with the data subsequently, including the validity of any research findings. In NaSH, there is the added problem of achieving a common level of minimum input across all health boards before research can potentially be done at the national level. Future research at this level will remain limited to variables included in the shared minimum dataset. However, there are many opportunities for ‘above and beyond’ research at local and regional level, and in theory all categorical data can be reported on via the anonymised data view, not just the agreed dataset variables.

Data left out of the collection template are 'lost' because the system is not designed in such a way for it to be found. This was repeatedly cited as a limitation in studies in our clinic literature sample. In many cases, the routinely collected data used did not include information on variables or behaviours that could have been valuable to the study, or that could be a source of residual confounding (see for instance Gindi et al. 2010). In a study from Seattle on circumcision and HIV, the authors noted that although data were available on the sexual roles of men in the sample (e.g. insertive, receptive or versatile) data were not available for the *frequency* of insertive and receptive sexual acts (Jameson et al. 2010). Similarly, a study by the Melbourne Sexual Health Centre on chlamydia positivity over time noted that data on sexual network characteristics (for example, concurrent relationships) were not collected during the study period, but could have been of interest (O'Rourke et al. 2009). This is partly a collection-design issue and partly an issue of data coverage. As noted earlier, NaSH is an extremely comprehensive database and can hold a large amount of data about very large populations over long periods of time. However, a routine clinical system has to function in a routine way and the data requirements must be balanced (and secondary to) the clinical needs.

Two clinics (one UK, one Australian) mentioned missing data as a specific issue. This is a reminder that the quality of base data is a fundamental consideration in using routine data for research. It is very difficult to go back and correct missing data after the fact, but it is very easy for data to be missed at collection stage through simple non-completion of fields. The Australian clinic noted that this was a particular issue at return visits, where there may be reluctance on the part of clinicians to ask again about sex worker status or history of injecting drug use, for example. Ways of overcoming problems with data collection that were cited included communicating with users to let them know why structured recording was important and educating staff about the importance of accuracy. This suggests that certain problems at the collection stage can be remedied by increasing staff awareness and engagement. The DSC Clinic noted that meetings with key users often helped to facilitate understanding of the data in the EPR system.

As a final point on data collection, several of the methodological papers dealing with the use of primary care data noted that reimbursement schemes can affect data collection (see for instance de Lusignan and van Weel 2006). There are reimbursement arrangements for clinical coding under the most recent GP contract in the UK, which covers Scottish GPs as well as those in England and Wales, with some variations, but these schemes do not apply to specialist sexual health settings.

### **3.10 Data storage**

Electronic data are sometimes cited as having fewer storage issues than paper records and they certainly take up little physical space. Ways of storing electronic data for research include data warehousing (Lau and Catchpole 2001) and cloud or grid computing (Flowers and Ferguson 2010). Warehousing refers to consolidating data from disparate databases and managing them within a single database. This brings resources together and facilitates analysis. Cloud or grid computing is where data are held on a distributed network of servers. Information can then be pulled down from wherever it sits, and in the required format. However, there are issues involved in storing electronic data: exceeding server capacity, or having to pay for storage at a data centre, for example. Hardware and software both need to be maintained and technical support provided.

In addition, because of the sensitive nature of much of the data in NaSH, there are obvious security issues around storing data anywhere other than the system itself, or transferring it out of the system. Within NaSH, data are secured through strictly controlled role-based access

arrangements. The data an individual can access are determined by their role, so that someone who is not directly involved in a patient's care cannot see that patient's full clinical record, for example. Data accessed by ISD for the minimum dataset and by the Board consortia for business reporting are anonymised. Should there be a move to extract a dataset for subsequent analysis there is always a danger that data could be lost or inappropriately accessed. This may undermine public and patient willingness to supply information in healthcare settings and erode confidence that their data will be looked after. The Patrick Clements Clinic noted that access to their system was restricted and any output was anonymised appropriately.

A related issue is that of data archiving in relation to storage time, rather than space. NHS records should be retained for seven years after the last patient contact or up to the age of 25 (if the patient is under 25 at last contact), before being deleted. The system has obviously been live for less than seven years, so this has never been put into practice. The seven-year mark in 2015 could be a flash point, requiring decisions about whether it is right to keep data beyond that point, how much can be retained and for how long, and feeding into considerations of ethics and confidentiality. A move to delete data at this point should also consider the public health and research benefits of keeping a de-identified anonymised data set around key social and sexual risk variables.

Another issue arising from NaSH itself is the ability of lifetime data to change. For example, a teenager with unsafe drinking recorded in 2011 may become an adult with safe drinking in 2016. Episode-based data remains true, but lifetime sexuality and smoking status, for example, can change over time, and the 'original' or preceding data are written over. Thus data are useful for cross-sectional analyses but less so for longitudinal studies. There is a need to develop a concept of archiving or storing key data items annually to 'freeze' them.

### **3.11 Data retrieval**

Data retrieval (being able to get data out of the system) was a common problem among the clinics we contacted. One UK clinic noted that data had to be on the system in a fixed, standardised format in order for it to be easily retrieved. This was undermined by a tendency for staff to use free text where they could, as this was perceived to be faster and was also more familiar from the previous use of paper notes. Consequently, it was important to identify the key pieces of information that had to be on the system in numeric or coded form to facilitate retrieval. This could be a particular issue for use of the NaSH anonymised data view because, as noted above, all free text items are removed. The DSC clinic noted that although their system was able to provide some baseline information and trend data, it would still be necessary to go into the individual records if more in-depth information was required for research analysis. In addition, once data had been retrieved, it could potentially be misinterpreted. The researcher needed to gain a good understanding of the system and of terminologies specific to the context. Although we have discussed how changes made to systems can improve data quality, such changes can also cause problems for the use of a dataset over time, as one clinic noted. It had changed the definition of the 'number of sexual partners' variable in its system from 'since last visit' to allow for the number in the last 3 months and the last 12 months to be recorded from 2009 onwards. This meant there were problems with using this variable in analyses with a timeframe extending before and beyond the change in definition.

The problem of time – having enough time, or the propensity of data retrieval and analysis to consume time – can be seen as a user issue, and one that links in with the problem of IT support, a

technical issue. Both of these were cited by different clinics. Above we referred to the American clinic which had a full-time member of staff for IT. The other American clinic, however, reported the opposite experience. Data management responsibilities were shared among several members of staff, including some who did not have data management training. Although the clinic posited that the easiest way to overcome this would be to hire more staff with the necessary expertise, this would be unlikely due to budgetary constraints. The necessity of having staff with good IT skills was also noted by the Patrick Clements Clinic in their review of EPR implementation in the clinic (Brooks et al. 2008).

Data extraction can be time consuming and complex. As noted by the DSC Clinic, obtaining the right combination of variables from the system in the first place can be difficult. Although several of the methodological studies discussed software, such as MIQUEST, that is used to run queries in computerised primary care data, there may still be training needs, financial resources required to purchase software, and compatibility or installation problems (see for example de Lusignan et al. 2006 and Majeed 2004). NaSH has its own proprietary query engine (AdHoc) but the underlying reporting database can be accessed by standard commercial query tools such as Crystal Reports and Business Objects, and a variety of these methods are already in use in several Boards. As reports run against a view of the reporting database, NaSH users have not faced issues of data extraction and manipulation.

### **3.12 Research governance**

There is a clear legal context around data sharing and confidentiality in the UK and there are particular arrangements specific to Scotland that would affect using data from NaSH. SHIP has conducted significant review of the research governance surrounding access to electronically recorded clinical data sets (Laurie and Syet, 2012), and the highly sensitive and personal nature of NaSH data suggests any research will require full review by the relevant Caldicott Guardian and Privacy Advisory Committee (PAC), following the high impact research pathway suggested by SHIP (Scottish Health Informatics Programme, 2012). Researchers outside the NHS could request data from the NaSH minimum dataset held by ISD, but would have to apply to, and meet the requirements of PAC. The use of data safe havens and systems for the secure linkage of data suggested by SHIP could provide means to enable access.

The Data Protection Act requires that organisations should use the minimum amount of data on a 'need to know' basis (data should be "adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed") and should retain information only for so long as it is required (Data Protection Act 1998). This last point in particular has implications for keeping NaSH data after a certain time period and for research access to that data for the purposes of longitudinal or retrospective studies. Furthermore, the existing anonymous data views were intended for use by skilled Health Board statistical analysts to support business planning and performance management. Wider use of NaSH data for secondary research will require formal review of the anonymisation algorithms.

The need to comply with national regulations for electronic records systems, address the security and confidentiality of data, and gain informed consent were noted as issues by some of the clinics we contacted, but most clinics did not report experiencing any problems. Baltimore stated that they had standing institutional review board approval to use their database for broad epidemiological research. Melbourne stated that patients were informed about the confidential handling of their information. In addition, when a patient did not wish their samples to be retained

and used for future research purposes, this was documented in the patient record and the samples were discarded. Only some Scottish clinics currently include a question to gain consent to contact patients for research purposes in NaSH.

Confidentiality is central to any use of sexual health data for wider research, and it has been argued that patient trust relating to the use of data needs to be earned (Chalmers and Muir 2006). A study looking at how patients at a GP practice in England felt about their medical data being shared on a new national database found that most were happy for their full records to be shared. However, there was less willingness to share data about issues that were perceived as sensitive or embarrassing, or that could potentially affect how the patients were viewed by others – chiefly mental health, sexual health and genito-urinary issues (Powell et al. 2006). A study conducted at Sydney Sexual Health Centre in 2009 found that patients were comfortable with their data being shared with other health care workers, but were increasingly less willing to have their data shared or used in ways that were not directly related to their healthcare (Ryder and McNulty 2009). In Scotland, SHIP has conducted public engagement focus groups on the collection, sharing and use of health data. Although participants were supportive of data sharing, this came with a number of caveats including desire for consent to be requested and obtained when data were to be used, even if they were to be anonymised (Scottish Health Informatics Programme 2011).

## 4. Conclusions and recommendations

Use of routinely collected data is a key, future priority for the Sexual Health programme at MRC/CSO SPHSU. Greater use and interrogation of such data could aid in the assessment of the indicators set out in the Scottish Government's Sexual Health and Blood Borne Virus Framework and have a key role in the future sexual health research agenda in Scotland (Scottish Government, 2011). It could give access to populations at high risk of poor sexual health outcomes, provide the means to respond to immediate research questions, and avoid initiating large-scale surveys unnecessarily. With a data set with over 700,000 patients and >200,000 attendances annually recorded, NaSH could have a clear place in future sexual health research.

In this review, we have reflected on the data collection, data management (including storage, IT, practicalities and quality assurance), data extraction and analysis, and data governance (including ethics, consent, access to data, and data sharing) issues requiring consideration in the use of NaSH data for sexual health research. It is clear that the quality of base data is a fundamental consideration in using routine data for research and, to facilitate data retrieval, it is imperative to identify the key information that should be on the system in numeric or coded form, as well as the user and technical proficiency required to maintain and access the data. There is potentially some conflict between the need for comprehensive and complete data for research purposes and the need for a routine clinical system to function in a routine way, within an acceptable timeframe and in a manner acceptable to patients and clinicians. Concerns over data collection, storage and retention should be considered within the context of the wider public health and research benefits of keeping a de-identified anonymised data set around key social and sexual risk variables.

To enable, and improve, use of NaSH data for sexual health research, we suggest the following recommendations. We also note the stakeholders who could act on the recommendations or help to ensure NaSH has a place in sexual health research in Scotland.

### 4.1 Continuation of NaSH

Future use of NaSH is currently under review by the Clinical Portfolio Management Group. Sexual health clinicians in Scotland record social and sexual risk assessments on tens of thousands of Scotland's population every year, many of them in harder to reach groups such as young men. Even though data completeness is likely patchy, there is clear merit in collecting these data and making further use of NaSH to explore sexual health in Scotland, for population-level monitoring of sexual risks, and to document trends in recorded risk behaviour over time. Learning from our review would be invaluable in ensuring further developments of NaSH, or any future replacement system, are correctly specified to be able to make maximum use of secondary data to inform social and sexual health research as well as public health and epidemiology.

**Action: Clinical Portfolio Management Group / NISG**

### 4.2 Recording willingness to be contacted for research

Only some clinics currently include a question to gain consent to contact patients for research purposes in NaSH. Given the strengths of NaSH in identifying subgroups of interest we recommend this question be included in the routine demographic set as a searchable data item, as well as incorporated in all registration forms and on-line registration if and when this happens. A

process of recording the consent acceptable to an ethics committee needs to be developed (e.g. by scanning on consent form). The process of being able to recruit potential research subjects is supported by the Caldicott review, which cites an example from the South London and Maudsley Trust (Caldicott et al, 2013, page 69). Researchers are given access to de-identified data in a 'safe haven', after which the EPR system administrator checks if the selected individuals have provided consent to be approached for research purposes and the details of those who have consented are released to the researcher at one of the partner organizations.

**Action: NaSH User Management Group/ NISG / AxSys/ Research Ethics Committee**

### **4.3 Data quality**

Implementation of NaSH has focused on clinical utility and improvements in clinic process. STISS coding was discontinued predicated on clinics recording a minimum dataset. However no formal process has been established to review data completeness or quality, although some boards have developed quite extensive local data quality reports. Until this is done, it is not possible to make clear recommendations that might improve the utility of routinely collected data for sexual health research. ISD urgently need to develop a series of data completeness checks now access to the national NaSH data view has Caldicott approval, mapping this to the expected agreed minimum data set. Simple tests for data integrity could also be run (e.g. looking for people with same-sex partners recorded in recent sexual history where the lifetime form fails to record this). The existing Lead Clinicians for each health board should be held accountable for oversight of data completeness and quality, and might consider directing some of the clerical and clinical resource saved through not doing coding towards improving data quality. Learning from other centres reviewed in this report includes closer monitoring of new staff and random notes audits (Sydney), and development of reminders and alerts over omissions (Melbourne). Boards should consider incorporating these into their data quality processes for NaSH.

**Action: ISD/HPS/ NaSH User Management Group/ Lead Clinicians**

### **4.4 Computer assisted completion / self-completion**

Few clinics internationally have adopted this as their primary method of history taking. NaSH users are developing models of paper-based self taken history along similarly cautious lines with clinicians rechecking the history and discussing the resulting risk assessment with the patient. Greater integration of kiosk-type or web-based forms should be considered; potentially allowing computer-based self completion. This would very likely improve data quality as the referential integrity checks that are missing in NaSH could be built into a CASI process. The NaSH User Group and Lead Clinicians should explore this and pilot in areas that are agreeable.

**Action: NaSH User Management Group/ NISG /AxSys/Lead Clinicians**

### **4.5 Data warehousing**

Data analysis in NaSH is complicated by the relational database and sheer scale of data items that may be drawn upon. To preserve data integrity due to time sensitive data handling we recommend key queries are run against the national reporting database and data extracted and stored in a simpler flat file format where the risk attribute as recorded at the time is tied to the outcome of interest, such as prescribing a specific drug or diagnosis of an STI (see Appendix 3). Setting up a secure data warehouse will require relevant permission and a formal data retention policy for NaSH.

A number of NaSH data items of social science research interest are stored in a 'single-record' form and can be amended at follow-up visits. Only the latest entry is written to the reporting database (although previous values can be seen if needed by the clinician). Examples include current alcohol and smoking use, and experience of gender-based violence. To mitigate against this, we recommend extracting these key data items into a protected data warehouse linking NaSH number with the values as known at a specific date, such as 31st December each year (see Appendix 3). This would allow identification of people with risk determinants in past years, some of which may now have moderated (eg alcohol or drug misuse). Only by storing an annual data set will it be possible to conduct true longitudinal studies looking at historic risk correlated to sexual health outcomes.

**Action: ISD/HPSI/NISG/NMAG/with advice from SHIP**

#### **4.6 Anonymisation process and relevance of data items**

The existing anonymous data views were intended for use by NHS Board statistical analysts to support business planning and performance management. Wider use of NaSH data for secondary research will require formal review of: the anonymisation algorithms; the degree to which additional data may be required; ..the possibilities for, and the implications of (including risk of deductive) disclosure; and the use of data safe havens and one-way linkage. Given the highly sensitive and personal nature of NaSH data any research involving NaSH data will require full review by the relevant Caldicott Guardian and Privacy Advisory Committee, and likely the high impact research pathway suggested by SHIP (Scottish Health Informatics Programme, 2012). The review should take into account developments in anonymisation since the data views were first established.

**Action: NaSH User Management Group/NMAG/Caldicott Guardian and Privacy Advisory Committee with responsibility for NaSH/with advice from SHIP.**

## 5. Acknowledgements

We are grateful to all those who provided assistance and made invaluable contributions to this review, in particular the clinics in the UK and abroad which provided information about their electronic systems.

Dr Steve Baguley *The City Clinic, Aberdeen*

Dr Gary Brook *Patrick Clements Clinic, London*

Professor Jackie Cassell *Division of Primary Care and Public Health, Brighton and Sussex Medical School, University of Brighton*

Dr James Chalmers *NHS Scotland Information Services Division*

Professor Christopher Fairley *Melbourne Sexual Health Centre*

Glenda Fehler *Melbourne Sexual Health Centre*

Candida Fenton *MRC Social and Public Health Sciences Unit*

Dr Khalil Ghanem *Baltimore City Health Department Public STD Clinics*

Dr Rachael Jones *West London Centre for Sexual Health, London*

George Laird *West of Scotland Managed Clinical Network*

Dr Roxanne Pieper Kerani *Harborview Medical Centre, Seattle*

Dr Kaveh Manavi *Whittall Street Clinic, Birmingham*

Dr Anna McNulty *Sydney Sexual Health Centre*

Dr Danielle Mercey *Mortimer Market Centre, London*

Dr Mark Pakianathan *The Courtyard Clinic, London*

Zareena Rafiq *NHS Scotland Information Services Division*

Mary Robins *MRC Social and Public Health Sciences Unit*

Dr Hiok-Hee Tan *DSC Clinic, Singapore*

## 6. References

- Brook, M.G. et al. (2008). Implementation of electronic patient records in a sexual health clinic. *Sexually Transmitted Infections* 84(2): 155-156.
- Caldicott F, et al. for the Department of Health (2013) Information: To share or not to share? The information governance review. London: Department of Health.
- Chalmers, J. and Muir, R. (2006) Patient privacy and confidentiality. *British Medical Journal* 326: 725-726.
- Data Protection Act (1998). (c29): London. Available online at <http://www.legislation.gov.uk/ukpga/1998/29>.
- de Lusignan, S. and van Weel, C. (2006). The use of routinely collected computer data for research in primary care: opportunities and challenges. *Family Practice* 23(2): 253-263.
- de Lusignan, S. et al. (2006b). Routinely collected general practice data: goldmines for research? A report of the European Federation for Medical Informatics Primary Care Informatics Working Group (EFMI PCIWG) from MIE2006, Maastricht, The Netherlands. *Informatics in Primary Care* 14(3): 203-209.
- Fairley, C. K., Sze, J. K., Vodstrcil, L. A., & Chen, M. Y. (2010) Computer-assisted interviewing in sexual health clinics. *Sexually Transmitted Diseases* 37:665-668.
- Flowers, J. and Ferguson, B. (2010). The future of health intelligence: challenges and opportunities. *Public Health* 124(5): 274-277.
- Ghanem K.G., et al. (2005). Audio computer assisted self interview and face to face interview modes in assessing response bias among STD clinic patients. *Sexually Transmitted Infections* 81(5): 421-425.
- Gindi, R.M. et al. (2010). Sexually Transmitted Infection Prevalence and Behavioral Risk Factors among Latino and Non-Latino Patients Attending the Baltimore City STD Clinics. *Sexually Transmitted Diseases* 37(3): 191–196.
- Information Commissioner's Office. (2012) Anonymisation: managing data protection risk code of practice. London: IOC.
- Jameson, D. et al. (2010). The association between lack of circumcision and HIV, HSV-2, and other sexually transmitted infections among men who have sex with men. *Sexually Transmitted Diseases* 37(3): 147-152.
- Lau, R. K. and Catchpole, M. (2001). Improving data collection and information retrieval for monitoring sexual health. *International Journal of STD & AIDS* 12(1): 8-13.

Laurie, G. And Syet, N. (2012) Information governance of use of health-related data in medical research in Scotland: Towards a good governance framework. University of Edinburgh, School of Law Research Paper Series, No. 2012/13. Edinburgh: University of Edinburgh.

Naughton et al. (2011). NaSH - National Sexual Health System Draft Paper for Clinical PMG. NHS Greater Glasgow & Clyde. Clinical Governance Snap Shot for Hub pan healthboard 01.01.11 – 31.12.11.

O'Rourke, K.M. et al. (2009). Trends in Chlamydia positivity over time among women in Melbourne Australia, 2003 to 2007. *Sexually Transmitted Diseases* 36(12): 763-767.

Paterson, N.I.A. for Scottish Government Health Department (2008b). Access Security Framework. Version 1.00.

Powell, J., Fitton, R. & Fitton, C. (2006) Sharing electronic health records: the patient view. *Informatics in Primary Care* 14:55-57.

Richens, J. et al. (2010). A randomised controlled trial of computer-assisted interviewing in sexual health clinics. *Sexually Transmitted Infections* 86(4): 310-314.

Ryder, N. et al. (2009). Increasing role of herpes simplex virus type 1 in first-episode anogenital herpes in heterosexual women and younger men who have sex with men, 1992-2006. *Sexually Transmitted Infections* 85(6): 416-419.

Scottish Executive (2005a). Respect and Responsibility: A Strategy and Action Plan for Improving Sexual Health. Edinburgh: Scottish Executive. Available online at <http://www.scotland.gov.uk/Topics/Health/health/sexualhealth/respect>.

Scottish Government (2011). Sexual Health and Blood Borne Virus Framework 2011-2015. Edinburgh: Scottish Government. Available online at <http://www.scotland.gov.uk/Publications/2011/08/24085708/0>.

Scottish Health Informatics Programme (2011) SHIP Public Engagement: Summary of Focus Group Findings. Version 1. Available online at [http://www.scot-ship.ac.uk/sites/default/files/Reports/Focus\\_Group\\_Findings\\_Briefing\\_Paper.pdf](http://www.scot-ship.ac.uk/sites/default/files/Reports/Focus_Group_Findings_Briefing_Paper.pdf).

Scottish Health Informatics Programme (2012) A Blueprint for Health Records Research In Scotland. Version 1. Available online at [http://www.scot-ship.ac.uk/sites/default/files/Reports/SHIP\\_BLUEPRINT\\_DOCUMENT\\_final\\_100712.pdf](http://www.scot-ship.ac.uk/sites/default/files/Reports/SHIP_BLUEPRINT_DOCUMENT_final_100712.pdf)

Thomas, R. And Walport, M. (2008) Data sharing review. Available at <http://www.connectingforhealth.nhs.uk/systemsandservices/infogov/links/datasharingreview.pdf>.

Tideman, R.L. et al. (2006). Client acceptability of the use of computers in a sexual health clinic. *International Journal of STD & AIDS* 17: 121-123.

Tideman, R. L. et al. (2007). 'A randomised controlled trial comparing computer-assisted with face-to-face sexual history taking in a clinical setting.' *Sexually Transmitted Infections* 83(1): 52-56.

Vodstrcil, L. A. et al. (2011) Computer assisted self interviewing in a sexual health clinic as part of routine clinical care: impact on service and patient and clinician views. PLoS One 6(3):e18456.

Winter, A.J. (2010). 'Managing Patients Differently: The Scottish National Sexual Health IT System (NaSH)' in Ilett, R. and Bigrigg, A. (eds) Transforming Sexual Health in Scotland: Cultural, Organisational and Partnership Approaches (e-book). Bentham Science Publishers: <http://www.benthamscience.com/ebooks/9781608050659/index.htm>.

## 7. Appendices

### **Appendix I: Clinic and parent department or hospital websites (sites accessed for data collection between November 2011 and May 2012; all links live at 06 September 2012)**

Courtyard Clinic: <http://www.courtyardclinic.nhs.uk>

Mortimer Market Centre:

<http://www.camdenproviderservices.nhs.uk/clinic/mortimer-market-centre>

Patrick Clements GUM Centre:

<http://www.pcch.demon.co.uk/Patrick.html> and

West London Centre for Sexual Health:

<http://www.chelwest.nhs.uk/services/hiv-sexual-health/clinics/west-london-centre-for-sexual-health/west-london-centre-for-sexual-health-wlchsh>

Whittall Street Clinic:

<http://www.whittallstreet.nhs.uk/>

Melbourne Sexual Health Centre:

<http://www.mshc.org.au>

Sydney Sexual Health Centre:

<http://www.sshc.org.au/>

Baltimore City Health Department STD Control Program/Public STD Clinics:

<http://baltimorehealth.org>

Public Health Seattle & King County STD Control Program/Public Health STD Clinic at Harborview Medical Centre:

<http://www.kingcounty.gov/healthservices/health.aspx> and <http://www.uwmedicine.org/patient-care/our-services/find-a-clinic/pages/clinic.aspx?clinicid=244>

DSC Clinic:

<http://www.dsc-sexualhealth.com.sg/> and <http://www.nsc.gov.sg/showpage.asp?id=33>

## Appendix 2: Year of system implementation, data collection methods, data collected, quality assurance and monitoring across clinics

CLINIC*	SYSTEM IMPLEMENTED	COLLECTION METHOD	DATA COLLECTED	QUALITY ASSURANCE AND MONITORING
<b>MMC</b>	2011	Real-time entry by clinician	Full EPR covering patient and sexual history	Compulsory fields for entry of key data items; coding accuracy checks via comparison with lab records or paired note review for free text; regular meetings to discuss system and evolving checks
<b>PCC</b>	2007	Real-time entry by clinician	Full history, examination and outcomes for GUM, contraception and HIV	Regular audits and discussions to improve quality
<b>WLCSH</b>	c.2000**	Real-time by clinician, lab results via pathology	Demographics; next of kin; GP details; lab results	–
<b>WSC</b>	2008	-	Most data essential for patient management	–
<b>MSHC</b>	2002 (paper and electronic) 2011 (fully paperless)	Patient self-completion combined with real-time check and entry by clinician, test results via labs	General patient information and demographics; sexual history and STI risk; medical history; outcomes; visit data	System evaluation pre-and post-implementation; CASI data confirmed by clinician during consultation; compulsory fields for entry of key data items; reminders and alerts for completion of fields; regular audits and reaction to feedback; training for users where required
<b>SSHC</b>	2001	Paper chart collection with later electronic entry, some patient self-completion via CASI	Administrative data; demographics; sexual/GUM history and risk data; lifestyle factors; outcomes; counselling	All files of new staff reviewed by manager to monitor error rate; random audits of clerking accuracy; random quality assurance of medical records
<b>BC</b>	1992 (scannable form) 2004 (real-time computerised system)	Real-time entry by clinician	Demographics; sexual risk profile; past medical history; physical examination details; outcomes	IT staff member responsible for integrity of data entry and storage; charts audit by medical directors

<b>SKC</b>	No EPR but chart data entered into electronic database since 1993	Paper chart collection with later electronic entry, some patient self-completion via CASI	Demographics; sexual history and risk behaviours; vaccination history; outcomes	No data monitoring plan currently in place but changes made at data collection stage to improve quality and reduce missing data.
<b>DSC</b>	2004	Real-time entry by staff, with data updated later as required.	Demographics; full clinical record including behavioural data; outcomes; counselling; contact tracing  Audit trail data, including users accessing the record; staff productivity indices and involvement in patient care management	Rigorous pre-implementation testing to ensure data quality.

\* MMC – Mortimer Market. PCC – Patrick Clements. WLC – West London. WSC – Whittall Street. MSHC – Melbourne. SSHC – Sydney. BC – Baltimore City. SKC – Seattle and King County.

\*\* Clinic responded that system was implemented “over ten years ago”.

**Appendix 3: Example of relevant data items that can be updated at subsequent visits**

<b>Nash Form</b>	<b>Field Name</b>
Lifetime Sexual History	Gender Previous Sexual Partners Sex with Overseas National Sexual Contact involving Payment Sex without consent
Blood Borne Virus Issues	Injecting Drug Status HIV test status Hepatitis B status Hepatitis B vaccination
Reproductive Health and Contraception History	Current Pregnancy Status  TOP/lifetime births/miscarriages Cervical smear test
Social History	Alcohol Drinking Status Smoking status Substance misuse Accommodation Type Domestic Abuse [GBV]
Demographics	Full postcode (link to area level deprivation / Health Board of residence)

Example of time-sensitive attributes that could be reported out per incidence of an event such as STI test outcome (for row-based flat file processing)

- Age at time of test / procedure / prescription
- SIMD quintile and Board of residence at time of query
- Location and type of clinic where test taken
- Same-sex experience
- Recreational including injecting drug use
- Alcohol use
- Smoking status
- GBV status
- Most recent recorded recent sexual history: new partners, number of partners

This could be output per calendar year against the following data sets:

- All Chlamydia tests
- All gonorrhoea NAAT tests
- All HIV tests
- LARC prescriptions
- Emergency Contraception forms



**MRC/CSO Social and Public Health Sciences Unit,  
University of Glasgow,  
4 Lilybank Gardens,  
Glasgow, G12 8RZ  
Tel: 0141 357 3949  
Fax: 0141 337 2389  
[www.sphsu.mrc.ac.uk](http://www.sphsu.mrc.ac.uk)**

A Research Unit supported by the Medical Research Council and the Chief Scientist Office of the Scottish Government Health Directorates, at the University of Glasgow.