



Scottish Information
Commissioner

**Decision 021/2005 Mr Michael Collie and the Common Services
Agency for the Scottish Health Service**

Request for incidences of childhood leukaemia in Dumfries and Galloway

**Applicant: Mr Michael Collie
Authority: Common Services Agency
Case No: 200500298
Decision Date: 15 August 2005**

**Kevin Dunion
Scottish Information Commissioner**

Kinburn Castle
Doubledykes Road
St Andrews
Fife
KY16 9DS



Decision 021/2005 – Mr Michael Collie and the Common Services Agency

Request for information relating to incidences of childhood leukaemia – information withheld on basis of section 38(1)(b) - failure of Common Services Agency to deal with request fully in accordance with section 1(1) of the Freedom of Information (Scotland) Act 2002.

Facts

Mr Collie requested information on incidences of childhood leukaemia for both sexes, in the age range 0 -14 by year from 1990 - 2003 for all of the Dumfries and Galloway (DG) postal area by census ward from the Common Services Agency (CSA).

Mr Collie was dissatisfied with the responses he received from the CSA to his initial request and to his subsequent request for review. Mr Collie lodged an application with the Commissioner to obtain the information he had requested.

Outcome

The Commissioner found that the Common Services Agency (the CSA) had not dealt with Mr Collie's request for information fully in accordance with Part 1 of the Freedom of Information (Scotland) Act 2002 (FOISA) in that it had breached section 1(1) of FOISA in not providing certain information by year at census ward level. The Commissioner required the authority to provide census ward data, suitably amended to protect against potential identification of individuals, or, if Mr Collie prefers, annual aggregate statistics at DG health board level.



Appeal

Should either the CSA or Mr Collie wish to appeal against this decision, there is an appeal to the Court of Session on a point of law only. Any such appeal must be made within 42 days of receipt of this notice.

Background

1. On 11 January 2005, Mr Collie requested information on incidences of childhood leukaemia for both sexes, in the age range 0 -14 by year from 1990 - 2003 for all of the Dumfries and Galloway (DG) postal area by census ward from the Common Services Agency for the Scottish Health Service (CSA).
2. The CSA responded to Mr Collie on 19 January 2005. It refused to release the information requested because the data for each of the years 1990-2001 by census ward resulted in small numbers of cases. As a result, it considered that this information fell within the definition of personal data under the terms of the Data Protection Act 1998 (the DPA 1998). The CSA further indicated that the data for the whole of the DG postal area for each year also resulted in very small number of cases and that therefore this information also fell within the definition of personal data. The CSA advised that it had a long standing rule of scrutinising any data containing small numbers before it is released and suppressing cells in tables containing less than 5 cases where it is considered that there is a significant risk of indirect identification of living individuals. This policy was applied by a number of organisations, similar to the CSA, which hold personal data. It indicated that the data for 2002 and 2003 was incomplete.
3. Mr Collie was dissatisfied with this response and on 19 January 2005 sought a review of this decision from the CSA. In its notice of review dated 26 January 2005 the CSA remained of the opinion that “these small numbers of cases at a local level carry a significant risk of disclosure of personal data on living individuals and are therefore exempt from disclosure as set out in FOI(S)A.” The CSA went on to state that “we are not suggesting that the data, in isolation, would necessarily identify an individual. But it is our understanding that, under the terms of the Data Protection Act 1998 we are required to regard data as personal if it might be possible to learn something about an individual from the data when combined with other information in the possession of, or likely to come into the possession of, the data recipient.”



4. The CSA acknowledged that it perceived the risks of disclosure to be low “but not absent” and stated that it was applying a “precautionary principle”. It felt this was justified given that “inadvertent disclosure could have adverse effects on the integrity of health information in Scotland if the CSA’s guardianship of the data was no longer trusted by the media and public.” It also pointed out that there was a potential alternative route to getting a question answered with much less risk of disclosure, by formally applying for research to be carried out into this area.
5. The CSA both responded to the initial request for information and carried out its internal review within the 20 working day time period specified in the Freedom of Information (Scotland) Act 2002 (FOISA). Its internal review appears, however, to have been carried out by the person who was responsible for the initial decision, contrary to the recommendations contained in the Code of Practice on the discharge of functions by public authorities under the Freedom of Information (Scotland) Act 2002 (Section 60 Code of Practice), although senior colleagues were consulted over the review decision. In its refusal notice, the CSA referred to the right of Mr Collie to appeal firstly to the CSA and then subsequently to the Scottish Information Commissioner. However, it did not reaffirm the right of appeal to the Commissioner in the subsequent notice of review, contrary to section 21(10) of FOISA. Finally, in its refusal notice and in its notice of review the CSA argued that the information was exempt under FOISA because the information was personal data in terms of the DPA 1998. The CSA did not actually cite the specific exemption as it appeared in FOISA, in this case section 38, or cite the specific subsection/paragraph. This is contrary to section 16(1) of FOISA.
6. The CSA has subsequently indicated that it has since reviewed and revised its policies and procedures and has undertaken further staff training on dealing with requests under FOISA.
7. The CSA offered to provide Mr Collie with data for the DG area for the combined period of 1990-2001.
8. I received an application from Michael Collie on 27 January 2005 requesting a decision following dissatisfaction with the notice from the CSA of 26 January 2005 refusing the release of information requested by him. The case was allocated to an Investigating Officer within my Office.

The Investigation

9. Mr Collie’s appeal was validated by establishing that he had made a valid information request to a Scottish public authority and had appealed to me only after requesting the authority to review its response to his request.



10. I subsequently invited comments from the CSA on the issues raised by this application and sought certain information from the CSA to assist with the investigation.
11. The CSA provided me with data for the whole DG area for the age range 0-14 for both sexes by year of diagnosis 1990-2001. This information confirmed the CSA's submissions that the numbers of cases in each year were all small. The CSA offered to provide the data broken down by census ward. Given that the primary issue for consideration in this investigation was whether an individual could be identified from the small numbers of cases I advised the CSA that I did not need to see the figures at census ward level for the purposes of the investigation as the total figures for each year for the whole DG area were very small.
12. In subsequent correspondence the data for the whole DG area were described by the CSA as registrations for the Dumfries and Galloway Health Board area. We sought clarification from the CSA on this issue.
13. The CSA explained that the DG postal area and the DG Health Board are virtually the same. There are 6 individual postcode units beginning with DG that are outwith the health board area. The average population of a full postcode unit in Scotland is about 35, but may be smaller in rural areas. As a result, the population of the postal area may be slightly larger.

The CSA's initial submissions

14. In its response to Mr Collie, the CSA referred to the DPA 1998 as grounds for withholding the information requested but did not refer explicitly to any exemption listed in Part II of FOISA. It subsequently confirmed to my Office that it was relying on section 38(1)(b) of FOISA.
15. Section 38(1)(b) exempts third party personal data in certain circumstances. During the investigation it became clear that the CSA was relying on section 38(1)(b) read in conjunction with section 38(2)(a)(i). Read together, these sections of FOISA exempt third party personal data from release if the release of the information would breach one of the data protection principles. Section 38(2)(a) refers to the definition of "personal data" contained in section 1(1) of the DPA 1998.
16. The CSA cited the seventh data protection principle as the principle that would be breached if the data was disclosed. The seventh principle states:

"Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data."



17. In its submissions to Mr Collie and to me, the CSA emphasised the need to achieve a balance between those who wish to see health information freely available and what it called “privacy enthusiasts”. It pointed out that there is no statutory requirement for NHS Boards to provide cancer data to the CSA and no obligation on patients to allow their data to be processed. The CSA is concerned that if the public loses confidence in the CSA, the CSA might not receive the data in the future.
18. In response to questions set by my Office, the CSA provided information about the supply and use of health data. National data sets are used for a variety of purposes. Although most are based on anonymised data, the CSA indicated that it is necessary to its operation to collect information that is identifiable. The CSA also indicated that there is no specific law underpinning the collection of data in this way. Explicit consent from patients is not obtained. Rather, the collection is undertaken on an implied consent basis. During the year 2000, cancer registries across the UK were faced with suspension of flows of data from some pathology laboratories and private sector companies following the General Medical Council’s (GMC) guidance which recommended doctors not submit data to cancer registries without explicit patient consent. The GMC modified its guidance temporarily.
19. The Health and Social Care Act 2001 which was subsequently introduced in England and Wales permits transfer of data until such time as arrangements are made for seeking and recording consent. In Scotland, these issues were addressed by the Confidentiality and Security Advisory Group for Scotland (CSAGS) which reported in 2002; *Report of the Confidentiality and Security Advisory Group for Scotland (CSAGS): Protecting patient confidentiality*. CSAGS recommended a strategy towards data collection on the basis of informed consent and/or “anonymisation”. However, CSAGS accepted that data collection could continue on the basis of implied consent providing patients were informed about the existence and uses of data and of their right to opt out of the system. Various patient and public information leaflets were developed to meet this “fair processing” requirement of the DPA 1998. The CSA provided me with a copy of a generic leaflet during the course of the investigation. The CSA contended that the uncertain legal status of national data sets leaves the CSA potentially vulnerable to continued attention of privacy activists.
20. In support of its refusal to supply the information requested to Mr Collie, the CSA argued that the small number of cases by year in each ward carries a significant risk of identifying a living individual. In particular, it indicated that the combination of rare diagnosis, specified age group and a small geographical area always raises the risk that an individual can be indirectly identified if other information is, or becomes, known.



21. In its submissions to me, the CSA set out its understanding of the DPA 1998. It indicated that it understood that under the terms of the DPA 1998 the CSA should not only consider the data being requested but is should also consider whether disclosure could arise because the data could be combined with other information known by the recipient or likely to come into his/her possession.

The CSA policy on cases of small numbers

22. In its submissions to my Office, the CSA sought to explain its policy on cases of small numbers. It supplied a number of documents in support of its submissions. Its own internal guidance states the following:

Paragraph 3.4 of ISD Confidentiality Rules for Staff - release of aggregated data

Certain aggregated statistics derived from individuals' personal data, because of the small numbers involved, enable a person to be identified. These cases must be discussed with an appropriate senior manager. If in doubt consult a Consultant in Public Health Medicine. Requests for this data from media or public should be cleared by a Consultant in Public Health Medicine.

23. The CSA indicated that, in practice, staff at the CSA regard numbers of five as a "cut off point" for consulting senior staff about data release. In its submissions, the CSA emphasised that it does not regard the cut-off numerator of less than five cases in any given cell of a statistical table as an absolute barrier to publication. Rather, it is a threshold which alerts the CSA to the need to review the table and come to a judgement as to whether there is a risk of contributing to the inadvertent disclosure of personal information about an individual.
24. I asked the CSA questions about the process followed by the CSA in each case to determine whether individuals could be identified from data. The CSA advised that there is "no hard and fast science to this, although there is a lot of practical experience." It is generally wary of tables:
- containing large numbers of cells with single cases and/or
 - with cells relating to small populations and/or
 - relating to small geographical areas.



25. The CSA submitted that in a relatively small population covering a small geographical area, each person's knowledge about other residents will obviously vary. Therefore, a child with leukaemia and the child's immediate family and friends are likely to be aware of the specific diagnosis and unlikely to learn anything new. But others who simply know that the child had cancer, could request an equivalent table for all childhood cancer combined and, assuming correspondence of the relevant data cell, would learn that the child had leukaemia.
26. The CSA supplied evidence from other organisations that have a similar policy on cases of small numbers. A letter from the Director of the Health and Care Division of the Office for National Statistics to the CSA dated 22 July 2002 stated the following:
- We advise that organisations such as yours should not release figures into the public domain that could potentially identify an individual. As an interim measure, while the risks are evaluated, we would advise that you do not present figures on the internet, NHS web or hardcopy public reports or publications that are based on a count of less than five.
27. The National Statistics Code of Practice *Protocol on Data Access and Confidentiality* (2004) states on page 7:
- The National Statistician will set standards for protecting confidentiality, including a guarantee that no statistics will be produced that are likely to identify an individual unless specifically agreed by them.
28. The UK Association of Cancer Registries (UKACR) *Guidelines on Release of a) individual level anonymised information and b) tabular information based on small populations or small cell counts (Version March 2005)* on release of potentially identifiable data, also supplied by the CSA, provide more specific guidance:
- Between these two extremes, the disclosure of anonymised but individual records or tabular data to a low level of aggregation (including some low cell counts) poses a very small but theoretical risk of identifiability when combined with other existing knowledge of the data recipient(s) or knowledge obtained from a different source.
29. The UKACR describes this kind of data as "potentially identifiable" because the extent of the other knowledge available to the recipient(s) of the data will usually be unknown. The guidelines go on to provide that the following categories of data should be regarded as "potentially identifiable:"
- tabular data based on small geographic areas, with cell counts of fewer than five cases/events



- tabular data containing cells that have underlying population denominators of less than approximately 1,000.
30. The UKACR guidance also contains guidelines on the publication of geographical data which states:
- Due to the rarity of cancer in children and young adults, there may be a non-negligible risk of information disclosure by publication in five year age groups between 0 and 24 years. In this age range, particular scrutiny should be paid to tabulations and appropriate aggregations should be made to avoid sparse cells.
- If the total denominator population of a geographical area is less than that of a Primary Care Trust, e.g. wards, release of tabular data based on such areas should conform with the above guidelines for the release of “potentially identifiable” data.
31. The CSA indicated that it is not bound by the UKACR guidance because it is primarily for registries in England but nonetheless supplied it to my Office because it “indicates current thinking on disclosure risks based on the interpretation of DPA 1998.”

Submissions from Mr Collie

32. In his application to me, Mr Collie indicated that he did not accept that the information requested in the detail sought could reasonably be held to risk identifying individuals in the localities concerned. He argued that the CSA deemed the risk to be “low” but had invoked the “precautionary principle”.
33. As part of his submissions, Mr Collie forwarded a report which had been provided to a local councillor in 2004. The report had been compiled by the CSA and looked at levels of childhood leukaemia in the area around Dundrennan in Dumfries and Galloway. The report looked at levels of all malignant neoplasms and, more specifically, leukaemias in adults as well as children ages 0-14. Levels of incidence of leukaemia in the Dundrennan postal code area were compared with levels of incidence at regional and national level.
34. The time period was 1992-2001 for the postcode sectors containing Dundrennan and the nearby firing range. For 0-14 age, childhood leukaemia rates were presented for both sexes combined. There were no cases for this time period so no cell suppression was required.
35. Mr Collie argued in his submissions to OSIC that the kind of information the CSA was withholding was essentially already in the public domain.



Research carried out by OSIC

36. In addition to questions to the CSA, I also undertook research into the issues raised by this case. This has involved looking at policies on cases of small data numbers in other jurisdictions, case law from other information commissioners and guidance from the Office of the Information Commissioner (IC) on the meaning of “personal data.” The IC is responsible for the regulation of DPA 1998 for the whole of the UK. In particular, I looked at:
- Washington State Department of Health: *Guidelines for working with small numbers*
 - European Network of Cancer Registries: *Guidelines on confidentiality in population-based cancer registration in the European Union*
 - Regina v Department of Health ex parte. Source Informatics Ltd
 - Information Commissioner: *Use and disclosure of health data* (2002)
37. On 17 June I sent the CSA an initial assessment which set out my preliminary views on this investigation and the issues that I considered to be unresolved. I raised a number of issues resulting from my own research and made particular reference to guidance such as the Washington State Department of Health guidelines (see above). I also queried the publication of certain data that currently appeared on the CSA’s website and how this related to this investigation. As a result of this initial assessment, the CSA made further submissions.

Final CSA submissions

38. The CSA made a number of final submissions to me, some of which I have addressed below in my analysis and findings. It did indicate, however, that it was much less concerned about the release of the data at Health Board level by year. It indicated that its real concern was the release of data at ward level. It accepted that the release of the data requested at Health Board level represents a much lower risk of disclosure.
39. As part of its final submissions the CSA forwarded a draft document it had developed for handling small numbers of data, *Draft Guidance on Handling Small Numbers* (this draft guidance has subsequently been published on the CSA’s website). This guidance would not prevent the CSA from releasing the data at Health Board level but would require it to perturb the data at ward level.
40. When a decision is made to perturb the data the CSA method employs a process known as “Barnardisation”. The method as employed by the CSA uses a modification rule: adding 0, +1, or -1 to all values where the true value lies in the range of 2 to 4 inclusive; adding 0 or +1 to cells where the value is 1; ‘0’s are kept as ‘0’.



Analysis and Findings

Discounted submissions

41. There are a number of submissions made by the CSA that I have discounted and I will deal with these first.
42. In its early submissions to me, the CSA cited only section 38(1)(b) as grounds for withholding the information requested in this case. (Although no exemption was specifically cited in its responses to Mr Collie, it is clear that the CSA were relying on this particular exemption.) In its final submission to me, however, the CSA cited section 36 (“confidentiality”) as an additional ground for withholding the information requested. It argued that while the CSA does not itself have a direct professional relationship with the patient, nevertheless, the relationship is part of a continuum between the patient, the GP and the organisation to which the GP, as part of his professional relationship, provides medical information. But for this relationship with the patient, the GP would not supply the CSA with the personal data.
43. The CSA did not indicate which part of section 36 applied and did not provide substantive arguments in support of its use. I have not considered the application of this exemption to the information withheld. I am of the view that consideration of this exemption would not affect my final decision as to whether the information should be released.
44. The CSA also indicated in its final submission to my Office that it had never taken an analysis of the DG area data by census ward and had stated this in its earlier correspondence with my Office. It indicated that as a consequence it did not have a copy of the data requested and that to create this would require “a significant bit of analytical work.” It indicated that it understood that it was only obliged to provide copies of data actually held and was not under an obligation to undertake new analyses in response to requests.
45. I am surprised that the CSA now states that to supply the data at census ward level would require a “significant bit of analytical work” as in correspondence to me dated 23 February it indicated that “this would be simple to undertake”.
46. I have looked at the CSA’s internal correspondence when Mr Collie’s request was first received. Following receipt of his request, a member of staff was asked to “pull the numbers”. The member of staff provided the figures in less than two hours for the whole DG area by year. Further correspondence indicated that if the data should be provided with the actual census wards that the cases occurred in then this could be “pulled” too.



47. In its correspondence with me, the CSA did not indicate that it was relying on this as grounds for not supplying the information to the applicant. This was also not cited as a ground when responding to Mr Collie's original request for information or his subsequent request for review. I have further noted that in its refusal notice to Mr Collie dated 19 January 2005 the CSA indicated that it did hold the data that Mr Collie requested for the time period 1990 to 2001 but not for 2002 or 2003.
48. As the CSA has not submitted explicitly that it is not supplying the information to Mr Collie in this case because it does not hold the information requested, I have discounted this issue in this case.
49. Further, the CSA raised a number of issues in its submissions to me which were not relevant to the application of section 38(1)(b). The CSA argued that the flow of data could be threatened if individuals no longer trusted the CSA to protect their personal health information. This argument is not relevant to the consideration of section 38(1)(b) the aspects of which are discussed below.
50. The CSA felt that emphasis should be given to its offer to Mr Collie to provide the information he wanted via an alternative route. That is, by requesting that specific research be carried out into this issue using well-established routes provided by the CSA. It had offered to carry out analysis within the CSA to address Mr Collie's concerns by supplying results such as rates for the aggregates of postcode sectors. I am not empowered to make any determination on this issue. Mr Collie's application to me relates solely to the CSA's refusal to provide him with the information under FOISA and that is the focus of this decision. Public authorities may make special arrangements with individuals and organisations to share information which would not be made publicly available. This is, however, outwith the scope of this application.
51. The CSA provided arguments for the public interest in its submissions to me. The public interest test only applies to section 38(1)(b) where the data subject(s) has/have issued a section 10 notice (right to prevent processing likely to cause damage or distress) under the DPA 1998. The CSA indicated that no patients had chosen to issue a section 10 notice. Therefore, even though there may be public interest grounds why this information should or should not be in the public domain, I cannot consider the public interest in this particular case.
52. I will now go on to consider the issues of substance.



Application of section 38(1)(b)

53. As mentioned above, section 38(2)(a) of FOISA refers explicitly to the definition of “personal data” contained in section 1(1) of the DPA 1998. “Personal data” are defined in section 1(1) of the DPA 1998 as follows:

data which relate to a living individual who can be identified:

- from those data
- from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.

54. It is clear that a diagnosis of leukaemia, if linked to an identifiable individual, is personal data under the terms of the DPA 1998. The only question in this case is whether a living individual can be identified from these data.

55. For the purposes of this investigation the issues for consideration were:

- Does the information requested constitute personal data?
- If yes, would release of this information breach one or more of the data protection principles?

56. I have proceeded on the basis that the individuals to whom this data relates are “living” because the CSA has not advised otherwise. If the individuals were no longer living, the DPA 1998 would, of course, not apply. Where the information related to a deceased individual who can be identified from that information the CSA might rely on section 38(1)(d) of FOISA which exempts a deceased person’s health record. The meaning of health record is that contained in section 1(1) of the Access to Health Records Act 1990 which states:

"health record" means a record which—

- (a) consists of information relating to the physical or mental health of an individual who can be identified from that information, or from that and other information in the possession of the holder of the record; and
- (b) has been made by or on behalf of a health professional in connection with the care of that individual

57. As can be seen from the definition above, to rely on section 38(1)(d), the CSA would also have to show that an individual can be identified from the data.



Does the information requested constitute “personal data”?

58. It is worth setting out at this stage the kind of information Mr Collie sought and what it would look like were it to be released. Mr Collie sought only numbers. He was not seeking any further information. I anticipate that the data would be provided in tabular form. Each box or cell within the table would denote a number which would correspond to the following information:
- age range 0-14
 - census ward
 - year of diagnosis.
59. The only information that could be obtained from this data is, for example,:
- That in year 1998 [number] of children in the age range 0-14 had been diagnosed with leukaemia in X census ward.
60. Were this information to be released at DG health board level the table would have a heading of age range 0-14 and then two columns; one listing year of diagnosis and the corresponding column listing a figure representing number of incidences.
61. Were this information to be released at census ward level, the table would contain 12 columns, one for each year of diagnosis (1990-2001), and 47 rows, one for each electoral ward in Dumfries and Galloway. There would be 564 cells containing a number. Some or many of these cells could contain zeros.
62. In its submissions to my Office the CSA did not focus on whether an individual could self-identify themselves from these data or whether a friend or relative familiar with the specific diagnosis could identify the subject of the data, although this clearly could be the case. Self-identification from statistics would be possible whether the cell contains figures of 1 or 500. Rather, the CSA's concerns have focussed on members of the public who already have a certain amount of knowledge about an individual, i.e. that they are ill, and that release of this data could provide the key to a specific diagnosis.



63. As a result, the CSA's submissions have emphasised that under the DPA 1998 it is required to consider not only the data being released but other information that might be in the possession of or come into the possession of the third party data recipient; in this case, information that might be in the possession of or come into the possession of any member of the public. Section 8(7) of the DPA 1998 provides that when considering requests for third party information the following should be considered:

For the purposes of section 7(4) and (5) another individual can be identified from the information being disclosed if he can be identified from that information, or from that and any other information which, in the reasonable belief of the data controller, is likely to be in or come into, the possession of the data subject making the request.

64. In the past, this section has been applied not only where a subject access request involves the release of third party data but also where any request for third party data is received. From 1 January 2005, however, all requests for third party data received by public authorities must be considered under FOISA and not DPA 1998. It was not clear whether section 8(7) could be taken into account when considering requests for third party data under FOISA given that section 38(2) refers only to the definition in section 1(1) of the DPA and guidance was sought from the Office of the Information Commissioner (OIC). The advice from the OIC was that section 8(7) of the DPA 1998 should be taken into account when considering whether an individual can be identified from data.

65. As a result of the application of section 8(7), I have therefore considered whether a living individual can be identified from those data or from that and any other information which, in the reasonable belief of the data controller, is likely to be in or come into, the possession of a member of the public.

66. In its submissions to Mr Collie and to me, the CSA pointed to its policy of cases where small numbers are involved and to similar policies from organisations with responsibilities for the collation and publication of statistics. The CSA submitted that its policy is an expression of its obligations under the DPA. Nevertheless, its policy reflects its own interpretation of the application of DPA.

67. From my own research, it is clear that concerns about the risk of identifying individuals when publishing cases of small numbers are shared by governments in other jurisdictions.



68. A number of cases before the Information and Privacy Commissioner in Ontario have involved figures and statistics being withheld on the grounds that disclosure would identify specific individuals. The Ministry of Health has relied on its own internal guidance which prohibits the release of “anonymised personal health information in tabulations of less than five in which a possibility exists where an individual person could be identified.” Likewise, the *Washington State Department of Health: Guidelines for working with small numbers* (the Washington guidelines) state the following:

In general, problems with confidentiality arise when there are small denominators (population size represented in a specific cell in a table); and, problems with data reliability arise when there are small numerators (cases in a specific cell in a table). In larger populations, it is more difficult to identify individuals from data released in tables.

69. I recognise that public authorities are concerned about the release of statistics which could potentially identify individuals, particularly where the data relates to sensitive information such as personal health data. The Washington guidelines are a useful example of overseas policy to address this particular problem.
70. However, reliance on a policy stance alone will be an insufficient basis for withholding information in response to a freedom of information request. The CSA indicates that its own internal guidance does not prohibit the release of small cell counts; rather it prompts consultation with a Consultant in Public Health to determine whether an individual could be identified from particular data.
71. It is clear from my research that simply because a cell contains small numbers, the cell is not automatically suppressed. For example, the Scottish Centre for Infection and Environmental Health, which uses the CSA portal, regularly publishes statistics on the incidence of HIV/AIDS. For the three month period from October 2004 to December 2004 a table sets out new reported incidences of HIV infection by health board and method of infection (SCIEH Weekly Report, p.176 Table 4). Many of the cells contain small numbers and in many cases numbers of 1 or 2. These cells have not been suppressed or “perturbed”.
72. The *Cancer Registration Statistics Scotland 1985-1995* (ISD: Cancer Registration Statistics Scotland 1986-1995 Edinburgh 1998) published on the CSA’s website includes data about specific cancers broken down by Health Board level, age and sex and contains many cells of small numbers. The updated statistics on the CSA’s website on incidences of specific cancers also contain cells of small numbers.
73. The key part of my investigation has been to try to ascertain the process followed by the CSA once a cell of numbers less than 5 has been identified. In particular, I have tried to identify the factors which will lead to a determination by the CSA that an individual could be identified from these data in any given case.



74. It is not clear from the above examples what criteria have been applied to the publications. In response to questions from my Office, the CSA has pointed to certain factors which would raise concerns that an individual could be identified from statistical information. These include tables:
- containing large numbers of cells with single cases and/or
 - with cells relating to small populations and/or
 - relating to small geographical areas.
75. In its submissions to my Office, however, the CSA did not expand on these factors. It did not define, for example, what it considered to be a small geographical area or a small population (although its subsequent draft guidance does address this.) I am uncomfortable with a policy based on such undefined criteria that could lead to information being withheld where the risk of identification is extremely low or non-existent. I have therefore tried to ascertain the kind of criteria that might apply in such cases that could lead to a finding that an individual could be identified.
76. I understand that these are difficult decisions to make and that applying strict criteria to each case may simplify what are necessarily complex decisions. However, for the sake of transparency, the CSA and other organisations holding this kind of data will need to be able to indicate the kind of factors which have been taken into consideration when deciding to withhold or to publish this type of information.
77. For example, it seems from my research that denominator (or population) size is a factor. It would be more difficult to argue that an individual could be identified from a single case in the whole of Scotland but less so where there is single case in a specific school.
78. During the course of the investigation, I have looked at the policies and practices in other jurisdictions to try and ascertain the criteria that might apply to such cases. The guidelines that I refer to in this decision are not a gold standard but rather illustrative of the kind of guidance that has been used in other jurisdictions.



79. For example, in the United States the Washington State Department of Health has developed guidelines on working with small numbers (see para. 35 above). The Washington guidelines provide that if there are 5,000 individuals in a specific age-race-sex group in a single county, the likelihood of identifying a single individual from data in a published table is quite small. However, it sets out a process for assessing possible disclosure of confidential information indicating that analysts should consider both denominator size and numerator size:

Analysts should first consider the size of the denominators ... Generally, tabular data based on denominators greater than 300 persons per cell present minimal risk for individual identification. The risk of violating confidentiality increases substantially when data are tabulated for small subgroups of the population within small geographic areas. Caution should be exercised by the analyst if the population size is between 100 and 300, and extreme caution is warranted when the population is less than 100.

Second, data analysts should consider the number of events in each cell of table to be released. If the count of cases or events in a cell is less than three, the data analysts needs to consider whether a breach of confidentiality is likely. A count of no events in the cell is clearly no threat to confidentiality, but a count of one or two events may be.

80. Census ward population figures for the DG postal code area for this age group are between 400-600 per ward. The combined population figure for the 47 census wards for this age group for the whole DG area and at health board level will be in the region of 19,000.
81. The CSA submitted that it is difficult to indicate the criteria that should apply in any given case. Nonetheless, in the course of its submissions to me it provided a copy of the guidelines of the UK Association of Cancer Registries. This does include some guidance on the kind of factors that should apply.
82. The UKACR's guidelines state at paragraph 6:

Provided that the total denominator population of a geographical area has approximately the size of a Primary Care Trust or greater, tabulations at the level of single sex, five-year age group (above 24 years), single year of incidence are permissible for publication in the public domain even if they contain a proportion of cells in which the denominator is less than 1000 and/or the numerator is less than five.



83. In relation to under 24 year olds it states:

Due to the rarity of cancer in children and young adults, there may be a non-negligible risk of information disclosure by publication in five year age groups between 0 and 24 years. In this age range, particular scrutiny should be paid to tabulations and appropriate aggregations should be made to avoid sparse cells.

If the total denominator population of a geographical area is less than that of a Primary Care Trust, e.g. wards, release of tabular data based on such areas should conform with the above guidelines for the release of “potentially identifiable” data.

84. Release of incidences of childhood leukaemia at health board level would not seem to be contrary to the Washington guidelines or the UKACR's guidelines. In the former case, while the number of incidences for a given year might be 1 or 2, the denominator size is significantly higher at around 19,000. In the latter case, while this case involves an age group under 24 years, the age range is much broader (0 -14).
85. In deciding whether individuals could be identified from this data (and from any other data that might be in the public domain) I have looked at the CSA's own practice in relation to the publication of statistics.
86. I have examined the data on incidences of cancer currently published on the CSA's website and have found that by combining information from adjacent tables significant information about an individual statistic can be deduced. For example, for each type of leukaemia, the CSA publishes two adjacent tables on incidence; one which provides statistics from 1975-2001 and another table providing a five year summary from 1997-2001. In both cases, the statistics are broken down by health board area. From combining information from both tables I could learn significant information about certain statistics.
87. In two examples, I knew the health board area, the type of leukaemia, the age range (within 5 years) and the year of diagnosis. Information that is currently published by the CSA can therefore be used in combination to produce quite detailed information about a specific statistic.
88. The CSA has argued that the combination of a rare diagnosis, specified age group and a small geographical area always raises the risk that an individual can be indirectly identified if other information is, or becomes, known. The CSA is itself publishing statistics of cells containing numbers of 1 or 2 where the diagnosis is a specific kind of leukaemia, the age group is within a 5 year range and at health board level.



89. The CSA has submitted that it is not possible to have consistency across an organisation of its size. However, I am simply trying to establish the criteria applied by the CSA in deciding whether an individual can be identified from this data.
90. The guidance I have considered during the course of the investigation and the CSA's own published statistics could lead to a view that this data could be released for each year requested at health board level.
91. However, Mr Collie sought the data at census ward level. Therefore, I have also considered whether a living individual can be identified from this data at ward level in the terms set out by section 8(7) of the DPA 1998.
92. In the guidelines I have examined, there is more concern about the potential for an individual to be identified where the data is published at census ward level. The UKACR guidelines, for example, state that if the total denominator population of a geographical area is less than that of a Primary Care Trust, e.g. wards or aggregation of wards, then release of tabular data based on such areas should conform with guidelines for the release of "potentially identifiable data".
93. An article on *Confidentiality and Data Access Issues for Institutional Review Boards* (George T Duncan: Confidentiality and Data Access Issues for Institutional Review Boards Protecting Participants and Facilitating Social and Behavioural Sciences Research (2003)) supplied by the CSA argues that data with particular characteristics pose substantial risk of disclosure and suggest vulnerability. It gives as examples:
- geographical details – census block
 - data with geographical details such as census tract data may be easily linked to known characteristic of respondents. Concern for this suggests placing minimum population levels for geographical identifiers.
94. I am also obliged to consider whether an individual can be identified not only from those types of data but from other data that may be in or could come into the public domain. The issue here, it seems to me, is not simply the population or denominator figure but also the geographical area. I consider that in a small geographical area, such as a census ward area, the residents will know more about each other. In particular, I consider it more likely that an individual may be aware that a child has had cancer but not know the specific diagnosis. The disclosure of this data could lead to a confirmation of a diagnosis of leukaemia.
95. As a result, I am satisfied that a living individual can be identified from this data at census ward level in terms of section 8(7) of the DPA 1998. Therefore I am satisfied that this information constitutes personal data as defined by section 1(1) of the DPA 1998.



Breach of the data protection principles

96. While I accept that the data at census ward level constitutes personal data, in order for the information to be exempt from disclosure, I still have to be satisfied that the release of the data would breach the data protection principles.
97. In my view the CSA was incorrect to argue that disclosure would breach the Seventh Data protection principle.
98. As previously mentioned, the seventh principle states:
- “Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.”
99. The CSA argued that if, as part of processing of the data, it shared this data in such a way as to disclose health information about an individual, then the CSA would have failed to take adequate steps to protect the data against unlawful processing.
100. I am not convinced that this is the appropriate principle to cite in this case. This principle deals with the type of security measures that an organisation, which processes personal data, should have in place to prevent, for example, unauthorised disclosure. The security systems in place in CSA would not be breached if this data were released. In any event, the CSA rules on confidentiality suggest that in cases of small numbers staff should consult a consultant in public health; the guidelines do not actually prohibit disclosure.
101. However, I am satisfied that disclosure of the information requested by Mr Collie would breach another of the data protection principles, i.e. the first principle which requires personal data to be processed fairly and lawfully.
102. Disclosure would be considered to be unlawful if, for example, the disclosure would constitute a breach of confidence. This could arise where sensitive information has been provided to a public authority in the expectation that it would not be disclosed.
103. I also consider the release of this information would be unfair. Guidance from OIC indicates that although the concept of “fairness” is harder to define than the concept of “lawfulness” the sorts of questions which should be considered include:
- Would the disclosure cause unnecessary or unjustified distress or damage to the person who the information is about?
 - Would the third party expect that his or her information might be disclosed to others?



- Had the person been led to believe that his or her information would be kept secret?
- Has the third party expressly refused consent to disclosure of the information?

104. I consider that a person would not expect their diagnosis of leukaemia to be placed in the public domain and would expect this information to remain confidential.
105. I am therefore satisfied that disclosure of the information sought by Mr Collie in its entirety would entail a breach of the first data protection principle and should not be released.
106. However this does not mean that Mr Collie should not have been provided with information and by failing to provide any of the information requested, the CSA did breach the provisions of section 1(1) FOISA.
107. I accept that the CSA offered to carry research into this area. In particular, it offered to carry out analysis within the CSA to address Mr Collie's concerns by supplying results such as rates for the aggregates of postcode sectors. However, the CSA could have provided certain information in response to Mr Collie's FOISA request.
108. Firstly, imagining as outlined above that census ward data would be made up of 564 cells (12 columns representing years, with 47 rows, one for each census ward), then some or many of those cells will contain zero. This does not contain personal information and so that information should have been provided to Mr Collie.
109. Secondly, where the cells which contained sparse numbers such that the release of the exact figure could identify an individual, then before withholding this information the CSA was under a duty to consider whether this information could be provided to Mr Collie in a less disclosive manner.
110. It could have done so by perturbing the data, that is by inserting a figure which is either greater, the same or less than the actual figure so that the risk of personal identification is substantially removed, provided that it had made it clear to Mr Collie that the figures were being perturbed and why .



111. As previously mentioned, the CSA latterly provided me with a copy of draft guidance that it has now published on its website. The guidance, entitled *Draft Guidance on Handling Small Numbers* sets out a process to be followed when handling the presentation of statistics where there is a potential risk disclosure of personal information as a result of small cell counts. The guidance considers both the denominator size and the numerator size. (The guidance also applies particular processes where the information is “Sensitive health-related data”. I note that, on the basis of this guidance, a diagnosis of leukaemia would not fall under this definition.)
112. Applying the methodology contained within that draft guidance would mean that Mr Collie could be provided with a table in which each cell was populated with a figure. Some or many of these would be zero. Others would be a ‘barnardised’ figure, that is perturbed by adding 0, +1, or -1 to all values where the true value lies in the range of 2 to 4 inclusive; adding 0 or +1 to cells where the value is 1.
113. This would provide the closest fit to fulfilling the request by Mr Collie for ‘details of all incidents of leukaemia for both sexes, in the age range 0-14 by year from 1990-2003 for all of the DG postal area by census ward’.
114. However, it does not provide all of the information he wants, as he could not be provided with the actual total by year or by period as that would make it likely that the actual figures for the perturbed data could be established.
115. So, under the duty to provide advice and assistance to Mr Collie under section 15(1) of FOISA, the CSA could have offered an alternative subset of information. This alternative would be to provide the actual (non barnardised) aggregate figures for the whole DG Health Board area for each of the years for which the information requested is held, as well as the total for all of those years. (This would also be consistent with its draft guidance which allows that where the figure for any one year may be less than 5 but greater than zero, that actual figure could be released because the geographical area of the health board is much greater than that of census ward to the extent that the possibility of identification is reasonably precluded.)



Decision

I find that the Common Services Agency (the CSA) did not deal with Mr Collie's request for information fully in accordance with Part 1 of the Freedom of Information (Scotland) Act 2002 (FOISA) in that it breached section 1(1) of FOISA in not providing certain information by year at census ward level for the Dumfries and Galloway postal area, for the years in which it held such information 1990-2001, as detailed above.

In respect of information for which an exemption applied, I find that the CSA did not provide advice and assistance to Mr Collie as to what information it was possible for it to supply to him as required under Section 15 of FOISA.

I require that the CSA provide Mr Collie with the census ward data for 1990-2001 for the DG postal area on the basis set out in paragraphs 112 to 114 above, that is, in a perturbed (barnardised) form unless Mr Collie would prefer to receive alternative information on aggregate annual figures for the whole DG Health Board area as indicated in paragraph 115 above.

I am obliged to give the CSA at least 42 days in which to supply Mr Collie with the information as set out above. In this case I require the CSA to enquire as to Mr Collie's preference and to provide him with the information within 2 months. For the avoidance of doubt if Mr Collie expresses no preference, then the CSA should provide Mr Collie with the census ward data for 1990-2001 for the DG postal area on the basis set out in paragraphs 112 to 114 above, that is, in a perturbed (barnardised) form within 2 months of the date of the receipt of this notice.

Finally, I find that the CSA breached Part 1 of FOISA in responding to Mr Collie's request for information as follows:

- section 16(1) – failure to specify the exemption being relied on to withhold information;



- section 21(10) – failure to advise Mr Collie of his rights of application to the Commissioner and of appeal conferred by FOISA.

However, I note that the CSA has subsequently indicated that it has since reviewed and revised its policies and procedures and has undertaken further staff training on dealing with requests under FOISA. As a result, I do not require the CSA to take any remedial steps in relation to the failures to comply with section 16(1) and 21(10) of FOISA.

Kevin Dunion
Scottish Information Commissioner
August 2005