THE NATIONAL HEALTH SERVICE NUMBER

R. D. T. FARMER AND K. W. CROSS

Royal College of General Practitioners Research Unit, Birmingham and Department of Social Medicine, University of Birmingham

Successful record linkage is dependent upon adequate identification of individuals. Acheson (1967) has stated the five fundamental criteria that should be fulfilled by an identification system: it should provide *unique, permanent* identification of *each* member of the population; it should be *available* on each of any pair of records to be linked, and it should be *economical*, i.e., it should consist of no more characters of information than necessary. He reviewed numbering systems currently in use, and one that fulfils three of his five criteria is the National Health Service number. It provides unique, permanent identification which covers virtually the whole population of the United Kingdom. It is no less available than any of the other allocated numbers in current use, and its availability could, perhaps, be increased by administrative changes without undue cost.

Although it is doubtful whether more than a small proportion of the population could memorize their National Health Service numbers, this disadvantage applies to any other allocated number. A number derived from nominal identifying data, such as the Hogben number (Hogben, Johnstone, and Cross, 1948), does not have this limitation; but Acheson (1967) has shown that discrepancies commonly occur in successive records derived by clerks from the same person when asked to give his or her surname (or maiden name), forenames, and date of birth.

Perhaps the principal drawback of the use of the National Health Service number as a means of record linkage is the heterogeneity of its structure. All National Health Service numbers consist of alpha and numeric characters. The alpha sequence is of variable length and is always followed by at least one numeric sequence, also of variable length. If there is more than one numeric sequence, they are separated by a slash (/); this is the only legitimate punctuation mark in the National Health Service numbers (excepting the Scottish birth registration number) and its correct location is an essential feature of the number. The numeric sequences are often separated by a full stop, a comma, or a dash; these should all be interpreted as a slash. Spaces

between the alpha and numeric sequences, or within either, have no significance.

Thus there are several series of numbers in use in England, Scotland, and Wales which collectively are referred to as the National Health Service number. Some of these series involve a sufficiently small number of people to make it feasible to consider renumbering them in line with one of the major series.

In view of the advantages, mentioned above, of the National Health Service number over other numbering systems currently in use, it is pertinent to examine these in detail.

HISTORICAL BACKGROUND

A National Register was compiled at the beginning of the second world war (September 1939) following an enumeration of the population of the United Kingdom on census principles. In the course of the compilation of that register a number was generated for each member of the population on the principles outlined below. Between September 1939 and February 1952, persons entering the country were allocated numbers and newborn children had numbers generated for them.

The main purpose of these numbers was the issue of unique identity cards and ration books for the whole population. For security reasons, persons who lost their identity cards after June 1940 were issued with new numbers on their new identity cards, and in such instances the old number and the new number were cross-referenced in the original register.

The system of national registration was abandoned in 1952. However, soon after the inception of the National Health Service in 1948, the numbers existing for national registration purposes were used as National Health Service numbers for administrative ease, thereby obviating the necessity to renumber the whole population. After 1952 numbers continued to be allocated to persons entering the United Kingdom and babies born within the country were given generated numbers.

A separate register was compiled, and is still maintained, for Scotland, but the number series are compatible with those for England and Wales, with one important exception which will be discussed later in this paper.

NUMBER SERIES CURRENTLY IN USE

PRIMARY NUMBERS (four-letter codes)

This series was generated during the enumeration of September 1939 when a number was allocated on the basis of where each person was staying on that night. The enumeration districts were identified by a four-letter sequence, and in Scotland the first alpha character within this sequence was always the letter S; in Northern Ireland the first alpha character was always U. Two numeric sequences were also used separated by a slash (/); the first identified the household in which the person was living at the time of the enumeration and the second specified the position of the individual within that household. The first numeric sequence rarely exceeds three digits (its maximum is four) but the second may occasionally consist of four digits. The large numbers sometimes attained in the latter numeric sequence are due to schools, hospitals, prisons, etc. being regarded as 'households'.

Thus the designation of this number is four alpha followed by two numeric sequences separated by a slash (/). Leading zeros are not used in this format.

<i>Examples</i>	QFHA	126/3
	KMAA	1/724
	SAAA	127/5 (for Scotland)
	UAAA	135/7 (for Northern Ireland)

PRIMARY NUMBERS (five-letter codes)

When the National Registration was initiated in 1939 it was found that some enumeration districts had larger populations than had been anticipated and sub-enumeration districts were created in these areas. The codes given to these sub-enumeration districts were derived by expanding the four-letter code for the appropriate enumeration district to a five-letter code, the fifth letter being added after the four-letter sequence. Otherwise the logic and format of this number series are exactly similar to those of the four-letter primary numbers.

<i>Examples</i>	EABPA	49/6
	LAPCA	16/67

Y NUMBERS

These numbers were allocated as replacements for numbers generated during the 1939 enumeration to persons who lost their identity cards between June 1940 and December 1946. These people were issued

with new identity cards and with new numbers. (Between September 1939 and June 1940 new identity cards were issued with the original number, and the character X was added after the last numeric sequence, e.g., QFHA 126/3X. The X is no longer used in these cases.)

The first letter of such a number was Y followed by the code designating the area in which the new identity card was issued, making a four-letter code. This alpha sequence was followed by a single numeric sequence of between one and seven digits. The same system was adopted in Scotland and Northern Ireland, the second alpha character being S in Scotland and U in Northern Ireland.

<i>Examples</i>	YABC	12345
	YTRT	123
	YSAB	12345 (for Scotland)
	YUAB	126 (for Northern Ireland)

NUMBERS ISSUED TO NEW ENTRANTS AND RE-ENTRANTS TO THE NATIONAL REGISTER

Persons entering the country after the National Register was compiled in 1939 until the ending of national registration in 1952, and persons released or discharged from the armed Forces were allocated numbers in order to obtain identity cards and ration books. The system that was adopted was a three alpha code designating the area in which the number was issued followed by a numeric series of six or seven digits. The numeric sequence may include leading zeros. This series is inconsistently written down inasmuch as the same number may be quoted with or without leading zeros (i.e., QAA 165493 and QAA 0165493 are two different ways of quoting the same number).

In Scotland, a three alpha code was used to designate the area in exactly the same way as in England, the first alpha character being the letter S, but the numeric series consisted of from one to six digits. Persons demobilized in Scotland were issued with a new number consisting of a three alpha sequence and a seven digit number.

<i>Examples</i>	ABC	123456
	TNN	1236795
	SCN	1695
	SDP	1234567 (for Scottish demobilization register)

OTHER NUMBERS ISSUED DURING EARLY PART OF THE WAR

Between 29 September 1939 and 12 August 1940, a person entering the country was issued with a number by the Immigration Department. The number comprised an alpha series of two or three

characters, the first letter of which was always P, and a single numeric series of between one and four digits.

Persons who were found homeless by the police at the time of national registration were allocated a number having the designation V followed by another alpha and a single numeric series.

Certain war refugees entering the country during the early part of the war were issued with a number beginning with the letters WR and followed by a numeric series of between three and six digits.

In Scotland, homeless people were issued with a number beginning with the letters SV and followed by a single numeric sequence; persons entering the country via a Scottish port were issued with a number SP followed by a single numeric sequence.

Less than half a million persons are covered by these systems.

Examples	PAA	67
	WR	1234
	SV	92
	SP	543

BIRTH REGISTER NUMBERS

All children born in this country after 29 September 1939 have been allocated numbers according to the registration areas in which they were born. Until the beginning of 1965 the number consists of a single alpha sequence of four characters, designating the registration area in which the child was born, followed by a single numeric sequence which cannot exceed 500. The alpha codes used in these numbers were not the same as those used to designate the districts in the 1939 census (primary numbers). Between 1939 and 1945 any four alpha combinations not used in the original registration, except those beginning with the letters P, V, U, S, I, and M, were allocated to birth registration districts. All possible combinations had been used by 1946 and birth registration districts were then given four-letter codes beginning with the letter M. These were exhausted by 1956 and districts were then allocated a four alpha code beginning with the letter P but excepting all combinations beginning with the letters PA. These combinations were exhausted in 1965 when the new birth codes were introduced (*vide infra*).

Example MAAA 26

This system applied only to England, Wales, Northern Ireland, and the Isle of Man; in Scotland a different system was used (*vide infra*).

NEW BIRTH CODE NUMBERS

In 1965 a new numbering system for the registration of births in England and Wales was introduced, which consisted of a five alpha sequence

followed by a single numeric sequence of between one and three digits. The first three alpha characters refer to the registration sub-district in which the child was originally registered, the fourth alpha character designates the year of registration, and the fifth alpha character designates the quarter in which the registration took place. If the first character in a five alpha sequence is the letter Z, the number is not a birth number (*vide infra*).

For children born in 1965 the fourth alpha character is A; B was used for 1966; C for 1967; D for 1968, etc. For children born in the March quarter, the fifth alpha character is M or N; J or K designates a birth in the June quarter; S or T designates a birth in the September quarter; and D or E a birth in the December quarter. Additional letters are sometimes used following the sequence outlined above except that the fourth alternative for use in the June quarter is the letter Z, in order to avoid duplication of the letter M.

When the fourth alpha character is A to G inclusive, the numeric sequence may not exceed 999. When the alpha terminator is H or a subsequent alpha the integer value of the numeric sequence may not exceed 500.

Examples WABAM 864
TIDBZ 12

This system is not used in Scotland.

NHS NUMBERS AND OTHER ALLOCATED NUMBERS

These numbers were introduced in February 1952. The series replaced the numbers issued to new entrants to the National Register. Between February 1952 and 1957, all persons entering the United Kingdom and not having a National Health Service number were issued with a number with the designation NHS followed by a single numeric sequence of up to seven digits.

Examples NHS 123456
NHS 169

Since 1957 persons born after September 1939 have been allocated a number similar to the birth registration numbers that were currently in use at the time of entry. Thus, between 1957 and 1965 the allocated number was a four alpha three numeric number beginning with the characters M or P. From 1965 until the present time the number allocated was of the five alpha three numeric type but the first alpha is the letter Z. In these circumstances the subsequent four alpha characters cease to be meaningful.

Examples ZABCD 123
ZCQPL 669

The designation NHS followed by a single numeric sequence continues to be allocated to

persons entering the country who were born before September 1939. The current maximum integer value is 4,186,830.

SCOTTISH BIRTH REGISTRATION NUMBERS

All children born in Scotland after 29 September 1939 were issued with a number which is completely different from any number used in England or Wales. The construction of this number has been described by Smith (1963), who also commented upon the problems arising from the inconsistency of form of National Health Service numbers.

The number begins with an S, followed by a sequence designating the registration district in which the child was born. Unfortunately, this sequence is not always wholly numeric and can include a *lower-case* alpha character in the fourth position, separated by a full stop from the preceding three characters. Otherwise the sequence consists of three digits, a full stop, and a further one or two digits. This is followed by a slash (/) and then the year of registration, which is written either in full or the last two digits only. A further slash separates the year of birth from a further numeric sequence of between one and four digits.

Examples S647.12/1965/23
S21.a/65/1

Scotland will be allocating an all-numeric birth registration number in the future.

As Smith points out, the problems of arranging such numbers intelligibly on a punch card are formidable, and we have been forced to exclude the number type from the taxonomy developed below for dealing with all other types of number. Fortunately, such numbers are rare in England; an analysis of the NHS numbers of 12,500 patients on the list of a group of general practitioners in Birmingham for which a computer-based registration system has been devised (Farmer and Cross, 1972) showed that this type of number accounted for only 0.4% of the practice population.

A FUNCTIONAL TAXONOMY OF NATIONAL HEALTH SERVICE NUMBERS

We have considered the development of the various series of the National Health Service number. The heterogeneity of the number is obvious. For functional purposes, we have classified the series as indicated in the Table, which also gives the format and approximate numbers of each type in use.

A total of 37 characters are used in this number, that is, 26 alpha, 10 numeric, and one punctuation mark (*viz* /). The spaces, where used when the number is written, are not significant. The maximum number of characters in any one number is 11. A number can be identified as being of a particular type by reference to the following discriminating points.

TABLE
(excluding Scottish birth registration numbers)

Number/Type	Date of Issue	Official Name	Format			Approximate Number in Current Use	Approximate Percentage of Population
			Alpha Series	First Numeric	Second Numeric		
IA	Sept. 1939	National Registration Numbers	AAAA	1-999	1-9999	16.5×10^6	31
IB	"	"	AAAAA	1-999	1-9999		
IIA	Sept. 1939— Feb. 1946	Birth Registration Numbers (1st series)	AAAA	1-500	Nil	19×10^6	37
	Feb. 1946— 56	"	MAAA	1-500	Nil		
IIB	1956-1965	"	PBAA	1-500	Nil	1×10^6	2
	June 1940- Dec. 1946	Y Numbers	YAAA	1-9999999	Nil		
III	1965-72	New Birth Code	AAAAA	1-999	Nil	5×10^6	10
	1972-	"	AAAAA	1-500	Nil		
IVA	Sept. 1939- end 1946	Current Register Numbers	AAA	111111-9999999	Nil	6×10^6	11
IVB	1952-	National Health Service Numbers	NHS	1-9999999	Nil	4.2×10^6	8
V	Other small series of numbers					Less than 1×10^6	1

1. All type I numbers have two numeric sequences separated by a slash (/).
2. All type II numbers have a single numeric series preceded by four alpha characters.
3. All type III numbers have a single numeric series preceded by five alpha characters.
4. All type IV numbers have a single numeric series preceded by three alpha characters.
5. All type V numbers have a single numeric sequence preceded by two alpha characters.

A two-stage matching process could be used for automatic manipulation of this number.

STAGE I

The number type is identified by means of an algorithm (an example of a possible algorithm is shown in the Appendix). The number is stored in alpha-numeric order within each series.

STAGE II

A match is made in the appropriate series by serial testing. As the discriminating points in each number type differ, a separate matching process is used in each number type.

CONCLUSION

The National Health Service number is available by reference to the individual's medical card, which should be in the person's possession. A person who loses his card can apply to the local National Health Service Executive Council for a duplicate which will, of course, carry the same number. The production of a medical card on each occasion that an individual uses the National Health Service is an educational problem which may be soluble by offering incentives to the population to carry a card having the number on their person, for example, a card similar to a banker's card or credit card.

With the exception of Scottish birth registration numbers, numbers may be handled automatically, using a two-stage matching process, although the effectiveness and efficiency of these procedures have yet to be demonstrated. However, preliminary experiments have given encouraging results. Furthermore, as many of the number series are logical, the number is of use for epidemiological investigations.

The authors gratefully acknowledge the help given by Mr. J. D. Smale of the National Health Service Central Register, Southport, and Mr. F. Rooke-Matthews and his staff of the General Register Office in supplying information relating to the number and providing helpful criticism of this approach. We also acknowledge the assistance given by Professor E. G. Knox of the Department of Social Medicine, University of Birmingham and his helpful comments and advice.

This work was undertaken while one of us (R.D.T.F.) held a Research Fellowship of the Royal College of General Practitioners funded by the Department of Health and Social Security.

REFERENCES

- ACHESON, E. D. (1967). *Medical Record Linkage*. Oxford University Press, London.
- FARMER, R. D. T., and CROSS, K. W. (1972). An automated records system for general practice. *Brit. J. prev. soc. Med.*, **26**, 148.
- HOGGEN, L., JOHNSTONE, M. M. L., and CROSS, K. W. (1948). Identification of medical documents. *Brit. med. J.*, **1**, 632.
- SMITH, ALWYN (1963). Automatic linkage of medical and vital registration records. *Brit. J. prev. soc. Med.*, **17**, 185.

Reprint requests to the Department of Social Medicine, The Medical School, Birmingham B15 2TJ.

APPENDIX

SUGGESTED ALGORITHM FOR VALIDATION OF NUMBER AND ALLOCATION TO TYPE

- begin* if there are any non-blank characters other than A-Z, 0-9, or / (slash) then NON-VALID; if more than one / (slash) then go to SCOTTISH BIRTH REGISTRATION NUMBER else go to ENGLISH;
- ENGLISH: if the first character is not alpha or the last character is not numeric or there are more than 5 alpha or there are more than 7 numeric or there are no alpha or there are no numeric or the first numeric is followed by any alpha or the character immediately preceding / is not 0-9 or the first character after / is not 1-9 then NON-VALID; else go to VALID.
- VALID: if any character is / then go to TYPE I else if 4 alpha characters then go to TYPE II else if 5 alpha characters then go to TYPE III else if 3 alpha characters then go to TYPE IV else if 2 alpha characters then go to TYPE V else NON-VALID;
- TYPE I: if there are 4 alpha characters then Type IA else if there are 5 alpha characters then Type IB else NON-VALID; FINISH;
- comment* More comprehensive validation of Type I is possible by reference to the list of 4 and 5 alpha codes used during the original enumeration and by testing for unused alpha sequences;
- TYPE II: if the first character is not Y and the numeric integer value less than 500 then Type IIA else if the first character is Y and the number of numeric characters is less than 7 then Type IIB else NON-VALID; FINISH;

comment The three alpha characters following the letter Y must be a permissible combination, and more comprehensive validation of Type II is possible by reference to codes issued during primary registration.

TYPE III: *if the numeric value is greater than 999 or last alpha is not any of M, N, O, P, J, K, L, Z, S, T, U, V, D, E, F, G, or the penultimate alpha character is not any of A, B, C, D, E, F, G, H, then NON-VALID else Type III; FINISH;*

comment Additional alpha characters will be accepted as valid in the penultimate alpha position after 1972. Note: J will be acceptable in 1973. More comprehensive validation of the first 3 alpha characters is also possible by reference to codes issued by registration districts and sub-districts;

TYPE IV: *if the first 3 alpha characters are NHS and there are not more than 7 numeric characters then Type IVB else if the first 3 alpha characters are not NHS and there are either 6 or 7 numerics then Type IVA else NON-VALID; FINISH;*

comment Again, more comprehensive validation is possible by checking for unused alpha sequences;

TYPE V: *if there are exactly 2 alpha characters then Type V else NON-VALID; FINISH;*

SCOTTISH BIRTH REGISTRATION NUMBER: Scottish

comment No further validation of this number is currently undertaken;

end

This validation routine forms part of the record system described by Farmer and Cross (1972).