

# Validation of two methods of long-term epidemiological follow-up

C C Patterson, E E McCrum, Z Mathewson, A E Evans

Accepted 18 June 1990.

---

## SUMMARY

*Two methods of long-term epidemiological follow-up were compared by using each to study the survival of 1622 myocardial infarction patients registered by the Belfast MONICA Project. Length of follow-up ranged between 3 and 5 years during which time 277 deaths were recorded.*

*A computer-based method for linking MONICA Project registration records with the Registrar General's death certification data identified 273 of the 277 deaths. Follow-up supplied by the Northern Ireland Central Services Agency through the flagging of patients in their master patient index identified 271 deaths; four of the six deaths which were missed occurred before computerisation of the index was complete. The study illustrates the value of computer-based linkage with death certification data and of flagging in the Central Services Agency master patient index.*

## INTRODUCTION

Obtaining long-term epidemiological follow-up on large numbers of patients can present substantial logistical problems. Nevertheless, Northern Ireland provides a good location for follow-up studies because of its position and its relatively low levels of migration in the older age-groups.

This paper compares the results obtained using two different methods to follow up patients registered by the Belfast MONICA Project. The first relied on a computerised search of death registration information collected by the Registrar General's Office, using personal identifying information recorded at death registration. The second was provided by the Central Services Agency through the flagging of patient records in the master patient index. Although the convenience of this approach is appealing, it is important that the method should be independently validated for completeness.

---

The WHO MONICA Project: Multinational MONitoring of trends and determinants in Cardiovascular diseases.

Department of Epidemiology and Public Health, The Queen's University of Belfast, Royal Victoria Hospital, Belfast BT12 6BJ.

C C Patterson, PhD, FSS, Lecturer in Medical Statistics.

E E McCrum, MSc, MIS, Research Fellow.

Z Mathewson, MD, MFCM, Medical Research Officer.

A E Evans, MD, FRCP, FFCM, Director, Belfast MONICA Project.

Correspondence to Dr Patterson.

## **MATERIALS AND METHODS**

### *The Belfast MONICA Project*

Since January 1983 the Belfast MONICA Project has sought to register every myocardial infarction occurring among individuals aged between 25 and 64 years and resident in the Belfast, Castlereagh, North Down or Ards district council areas. During the period 1983–85 a total of 2,727 events which fulfilled the Project criteria for definite or possible infarction<sup>1</sup> were registered. These events occurred among 2,512 residents, 1,622 of whom survived at least 28 days after onset of the infarction. Long-term follow-up of these 1,622 individuals was obtained to 30th June 1988.

### *Record Linkage*

At the start of the Belfast MONICA Project it was appreciated that the follow-up of such large numbers of patients would be a major undertaking. It was therefore decided to automate follow-up by routinely linking computerised MONICA records with the Registrar General's computerised death registration records. Deaths registered in all age-groups throughout Northern Ireland during the years 1983–88 were considered. Patients who left the Project area to live in another part of Northern Ireland were therefore not lost to follow-up.

As with all large-scale data collection exercises, errors in the information recorded may result for a variety of reasons. As well as errors in data coding and preparation, misreporting errors can occur. Much of the information recorded at death registration is supplied by an informant who may not necessarily have been related to the deceased. Special steps were taken to ensure that errors in recorded data did not result in a failure of the search procedure to link MONICA records with matching death registrations. The approach built on the experience of the Northern Ireland Record Linkage Research Unit,<sup>2</sup> but used a rather different methodology.<sup>3</sup>

Ideally the personal identifying information available for such a linkage exercise should be permanent and have high discriminating power. Although the National Health Service number comes closest to this ideal, it is seldom known and is not recorded either by the MONICA Project or at death registration. Surname, marital status and occupation all lack permanence. Although forenames are less likely to change, they are often reported inaccurately with the use of abbreviated forms or inversion of order common. Surnames are sometimes misspelt with similar versions of the same surname confused. To minimise the effects of this problem the "Russell Soundex code" was used. Similar versions of the same surname all have the same Soundex code (eg Smith, Smyth, Smythe), so use of the code in place of the surname in the search procedure can deal with the majority of discrepancies in the spelling of surnames.<sup>4</sup>

Whether or not the search was successful was decided on the basis of "weights of comparison" derived from the Soundex code, the forenames and initials, the day, month and year of birth and the district council area of residence. The computer used these weights to mimic the intuition of the human mind. For example, one would be more likely to match two records which agree on a rare surname such as "Gravenitz" than two records which agree on a common surname such as "Smith". An illustration of the calculation of weights is shown in Fig 1. Although the first pair of records receive positive weights for agreement on surname and near-agreement on the forename and year of birth, the disagreement on the day

CALCULATION OF WEIGHTS OF COMPARISON						
<i>Soundex code and surname</i>	<i>Forename</i>	<i>Date of birth</i>			<i>District Council</i>	
B650 BROWN	FRANK	04 JUL 1930			BELFAST	
B650 BROWN	FRANCIS	10 OCT 1929			N. DOWN	
+9	+3	-3	-4	+2	-3	▪ 4
Complete agreement	Near agreement	Near agreement on year			Disagreement	
C462 CLARKE	ANNIE J	23 JUN 1918			ARDS	
C462 CLARK	JANE	23 AUG 1918			ARDS	
+5	-1	+5	-4	+7	+4	▪ 16
Only Soundex agrees	Agree on initial	Agree on day and year			Agreement	

Fig 1. Use of identifying information in patient records to derive a weight of comparison for assessing the likelihood that two records relate to the same individual.

and month of birth and the district council are sufficient to give a low total weight of 4. In contrast the second pair of records give a high total weight of 16 despite discrepancies in forename, surname and month of birth. The derivation of the weights has been fully described elsewhere.<sup>5</sup>

Weights for a sample of 116 matching MONICA and death registration records and for a sample of 5,345 non-matching records are shown in Fig 2. The complete separation of the two distributions by the selected cut-off of 10 units illustrates the ability of the search procedure to distinguish between matching and non-matching records. The patient's sex, marital status, occupation and address were used to verify manually all matches generated by the search procedure.

#### *Central Services Agency flagging*

In 1983 the Northern Ireland Central Services Agency began to computerise its master patient index, so permitting the flagging of records of individual patients. The index is updated primarily by using data from death certification, but also by using information received from general practitioners and from relatives returning medical cards to the Agency. Patient transfers to health authorities in Great Britain are recorded. By 1986 the Agency was in a position to supply follow-up information on patients from mid 1983. A similar system has operated successfully in Great Britain for many years through the National Health Service Central Register.<sup>6</sup>

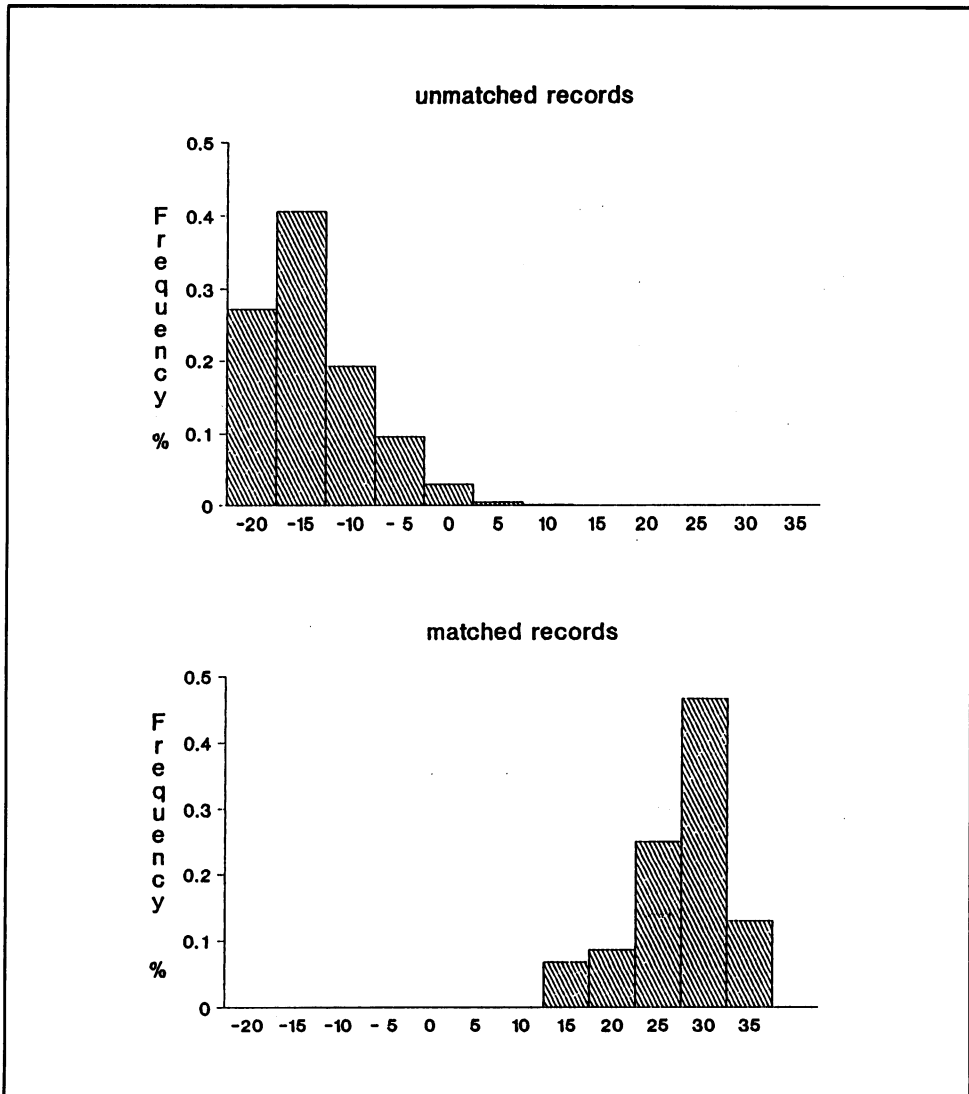


Fig 2. The distribution of weights of comparison in 116 pairs of records from the same individual (matched) and in 5,345 pairs of records from different individuals (unmatched).

## RESULTS

A comparison of follow-up results obtained by the two methods is shown in the Table. A total of 22 patients could not be flagged because no entry with matching details could be found in the master patient index. Ten patients were notified as having transferred to health authorities in Great Britain, and were considered lost to follow-up.

Four deaths were missed by record linkage. One occurred in England and was registered there. Another produced a weight of only 8 because of a difference in forename (Eithne instead of Ann), a missing day and month of birth on the

TABLE

*Comparison of Central Services Agency and Record Linkage follow-up for 1622 myocardial infarction survivors*

<i>Central Services Agency follow-up</i>	<i>Record Linkage follow-up</i>	
	<i>Dead (matched)</i>	<i>Alive (unmatched)</i>
Dead	267	4
Alive	2	1317
Lost to follow-up	0	10
Not flagged	4	18

MONICA record and a discrepancy in the year of birth (1928 instead of 1929). The remaining two deaths were confirmed by the patient's general practitioner, but death certifications could not be traced despite extensive searches. Two deaths identified by record linkage were overlooked by Central Services Agency flagging, and further checks suggested that these were probably clerical errors by the Agency. The four deaths occurring among individuals who could not be flagged occurred in early 1983 before computerisation of the Agency's index was completed.

As a further check on the completeness of follow-up, 277 general practitioners were mailed in July 1988 with a request for follow-up information on the 1622 patients. Follow-up information was obtained for 1,385 (85%) of the patients. Nine (4%) of the 223 deaths notified had apparently been missed by record linkage and flagging. Two auxiliary sources were used to check the status of these nine patients. The first was hospital records, and these confirmed that three of the patients had been reviewed after 30th June 1988. The second was the Northern Ireland 1988 electoral register, the qualifying date for which was 15th September 1988. Of the remaining six patients, four were still on the 1988 register and were therefore assumed to be alive at the end of follow-up. One of the final two patients had previously been reported by his general practitioner as being alive in July 1987, but the other could not be traced at all.

## DISCUSSION

The flagging of patients in the Central Services Agency master patient index performed very satisfactorily missing only six deaths in the follow-up period, four of which occurred before computerisation of the index was complete. Any chance of clerical error has been reduced by the recent introduction of computer-generated follow-up reports. The attractions of flagging are its convenience and completeness, and its ability to identify some patients who emigrate. The flagging facility will prove extremely valuable to those involved in epidemiological research in the Province. Additionally, at a time of increasing interest in clinical audit, there will be an important role for this facility in the evaluation of the long-term outcome of medical care. The introduction of this service by the Central Services Agency is therefore both welcome and timely.

The computerised search procedure developed to link MONICA Project records with death certification data was also found to perform well, ascertaining all but

four of the 277 deaths identified in the follow-up period. This procedure does retain some advantage over flagging. It may be performed retrospectively, while the flagging has only a limited retrospective capability. Record linkage also directly identifies the Registrar General's serial number, thus simplifying ascertainment of the registered cause of death.

The broader issue of the difficulty of linking computerised records from different sources (general practice sessions, inpatient admissions, outpatient attendances, screening and immunisation clinics etc) has yet to be resolved. Only when a unique identification number is in widespread general use can the information technology revolution be expected to make its fullest contribution to research in the health service.

The authors acknowledge the kind co-operation of the staff of the Registrar General's Office, the Central Services Agency and the general practitioners of the study area without whose assistance this work would not have been possible.

#### REFERENCES

1. World Health Organisation MONICA Project Principal Investigators. The World Health Organisation MONICA Project: a major international collaboration. *J Clin Epidemiol* 1988; **41**: 105-14.
2. Butler AR, Kar Ng WT, McDonald JR. Procedures for making linkage decisions in the Northern Ireland computer-based system of linked medical records. Northern Ireland Medical Record Linkage Research Unit, Belfast, 1972.
3. Howe GR, Lindsay J. A generalised iterative record linkage computer system for use in medical follow-up studies. *Comp Biomed Res* 1981; **14**: 327-40.
4. Newcombe HB, Kennedy JM. Record linkage: making maximum use of the discriminating power of identifying information. *Comm Assoc Comp Mach* 1962; **5**: 563-6.
5. Patterson CC. Survivorship models applied to the study of coronary heart disease. Unpublished PhD dissertation. The Queen's University of Belfast, 1989.
6. Baldwin JA, Acheson ED, Graham WJ, eds. Textbook of medical record linkage. Oxford: Oxford University Press, 1987.