**QUEEN'S
UNIVERSITY
BELFAST**

The epidemiology of motor neuron disease in Northern Ireland using capture recapture methodology

Donaghy, C. G., Clarke, J., Patterson, C., Kee, F., Hardiman, O., & Patterson, V. (2010). The epidemiology of motor neuron disease in Northern Ireland using capture recapture methodology. *Amyotrophic Lateral Sclerosis*, 11(4), 374-378. <https://doi.org/10.3109/17482960903329569>

Published in:
Amyotrophic Lateral Sclerosis

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

ORIGINAL ARTICLE

The epidemiology of motor neuron disease in Northern Ireland using capture-recapture methodology

COLETTE DONAGHY¹, JUDITH CLARKE², CHRIS PATTERSON³, FRANK KEE⁴,
ORLA HARDIMAN⁵ & VICTOR PATTERSON¹

¹Department of Neurology, ²Clinical Research Support Centre, ⁴Department of Epidemiology and Public Health, Royal Victoria Hospital, Belfast, ³School of Medicine, Dentistry and Biomedical Sciences, Queens University Royal Victoria Hospital, Belfast, Northern Ireland, and ⁵Department of Neurology, Beaumont Hospital, Dublin, Ireland

Abstract

Capture-recapture analysis allows an investigator to estimate the number of unobserved cases in an epidemiological survey, although it has not been employed widely in published research to date. This study examines the usefulness of capture-recapture analysis in measuring the incidence and prevalence of motor neuron disease (MND) in Northern Ireland (NI). Cases were ascertained from the NI MND register to measure incidence and prevalence. The authors then employed capture-recapture analysis to estimate the number of missing cases. Less than one case was estimated to be missing from the prevalence study, providing a completeness of ascertainment of 99%. Prevalence was 3.3 per 100,000 population adjusted to the European standard population and was unchanged when adjusted for missing cases. Incidence was 1.4 per 100,000 person-years adjusted to the European standard population. Due to the lack of overlapping sources it was inferred that there were no missing cases and completeness of ascertainment was estimated at 100%. In conclusion, capture-recapture analysis is relatively simple to perform and provides an objective measure of case ascertainment in epidemiological studies. The authors feel that future investigators should consider employing capture-recapture methods, where possible, to validate their estimations of disease frequency in MND.

Key words: Capture-recapture, motor neuron disease, prevalence, incidence

Introduction

In any epidemiological study of human disease, it is never certain how many cases remain unobserved. However, it is important to estimate the number of unobserved cases rather than making an unrealistic assumption of complete ascertainment. Capture-recapture analysis can help in this regard and has four main applications for use in human epidemiology. It allows an investigator to estimate the number of cases in a population even if only incomplete sources are available, to check the completeness of a registry, to refine disease rates in a population, and to provide lower and upper estimates of the total number of cases within a population (1).

Few epidemiological studies of MND to date have employed capture-recapture methodology (2–6). The aim of this study was to measure the incidence

and prevalence of MND in NI and to highlight the usefulness of capture-recapture methods in refining the disease rates and providing an estimate of the completeness of ascertainment.

Methods

Calculation of crude rates

Prevalence day for this survey was taken as 30 June 2005 and incidence was measured from 1 August 2004 to 31 July 2006. For the calculation of incidence and prevalence rates the 2004 mid-year estimate for the Northern Ireland population of 1,710,322 was used as the denominator (7) and 95% confidence limits for rates were calculated using Poisson distribution. Rates were then adjusted to the European standard population (8).

Correspondence: C. Donaghy, Department of Neurology, Royal Victoria Hospital, Belfast BT12 6BA, N. Ireland. Tel: 02890 634325. E-mail: donaghyla@hotmail.com

(Received 6 June 2009; accepted 10 June 2009)

ISSN 1748-2968 print/ISSN 1471-180X online © 2010 Informa UK Ltd. (Informa Healthcare, Taylor & Francis AS)
DOI: 10.3109/17482960903329569

Capture-recapture methods

Capture-recapture analysis is so named because of its origin in ecology studies that employ marking experiments designed to estimate the size of an animal population; it is also known as the 'Petersen method' (9,10). A widely-used example of the capture-recapture approach describes a method to estimate the size of a pool of fish involving first the capture of a net of fish, and then the subsequent marking of them before release. A second net of fish would then be captured and the proportion of fish marked would allow the probable size of the entire pool to be estimated. In human epidemiology the basic method of capture-recapture analysis requires an estimation to be made of the number of unobserved cases by analysing the degree of overlap between sources of cases. In practice, this means that each case found must have been logged beside its source(s) from which it was identified to allow the investigator to trace their intersection. This is critical in the planning of any study intending to employ capture-recapture methods. The transition into human epidemiology was first demonstrated by Wittes *et al.* (11,12) in the 1960s who highlighted some of the difficulties in applying this methodology to human populations. When a marked sample of animals (the 'capture') is released back into the population from which it came, the animals are generally able to redistribute, allowing the marked cases to mix well with the remaining population. This then allows the second sample (the 'recapture') to be independent of the first. Sources within human epidemiology are rarely entirely independent of each other, but dependencies between sources may be examined and accounted for using capture-recapture analysis. Another important consideration is that all cases should have equal 'catchability', which, in human studies, means that certain cases should not be more visible to sources than others. Furthermore, the population should be closed, which is unlikely in many human disease populations due to death and migration. These are the main challenges to be faced when applying capture-recapture methods to human epidemiology. These form the basis of the explicit 'assumptions' that need to be considered when employing this methodology, while implicit assumptions require all cases to be true cases and be recorded accurately (1).

Case selection and documentation

Cases were ascertained from the NI MND register and all persons gave informed consent prior to their inclusion in the study. The study protocol was approved by the local ethics committee (Office for Research Ethics Committees in Northern Ireland, ORECNI) and was performed in accordance with the ethical standards laid down in the Declaration of Helsinki. Six sources were utilized for case ascertainment: 1) the neurology department (involving a computer search of all outpatient clinic letters and

inpatient discharges); 2) the MND association members list; 3) general practitioners (a letter was sent to each practice in NI); 4) the regional pharmacy unit (this unit holds the names of all patients on riluzole in NI); 5) hospital coding lists from all acute hospital trusts in NI; and 6) electromyography (EMG) records from the three neurophysiology departments in NI. A thorough search was then undertaken within each source and within the appropriate time-frame to find as many cases as possible. Potential cases were validated by one of the authors (CD) before inclusion into the study. This was achieved first by an examination of case notes and secondly by a clinical consultation. For each case, a note was then made as to the source(s) by which that case was identified. Personal data, including hospital numbers, served as a 'unique identifier' so that a patient found in a second source could be cross-referenced to the first. Each record was stored in the NI MND register. Using an access database, the numbers of cases within each combination of sources was calculated for both the incidence and prevalence study. No cases were identified by general practitioners or EMG records that were not already ascertained via other sources and so these sources were excluded from the analysis. This resulted in a four-source capture-recapture analysis.

Capture-recapture analysis

Where two sources are available in an epidemiological survey a two-source analysis is performed. This is the simplest type of analysis and can be likened to the two-by-two contingency table used in calculating an odds ratio but with one entry missing. This analysis assumes independence between the groups and cannot quantify dependencies, although sources known to be positively dependent will tend to underestimate population size while negative dependency tends to overestimate it (1). Where more than two sources are available, log-linear modelling (1) can be employed to decipher the likely dependencies between the sources by examining all possible two-way interactions. A three- or four-source analysis is most practical as any more would incur unmanageable numbers of possible interactions. Where multiple sources exist, collapsing small or positively dependent sources together can help create a more manageable three-source analysis. In this study we employed a four-source analysis. Statistical packages (e.g. STATA or GLIM) will fit various models to data and will provide a goodness-of-fit χ^2 statistic, G^2 , for every model with a lower value of G^2 indicating better fit. For our four-source analysis we examined all two-way interactions separately and in combination, providing a model for each. In general, the simplest model that provides no evidence of lack of fit to the data should be chosen, in keeping with the principle of parsimony, as needlessly complex models may provide more unstable and unreliable

Table I. Data for four-source capture-recapture. All known prevalent MND cases are stratified according to the source(s) from which they can be identified. These are the basic data required for a statistical package to perform log-linear modelling. The cell denoted by '?' signifies the missing cases that have not been identified from any of the sources. Estimating the number of unascertained or missing cases is the purpose of the capture-recapture analysis.

				Neurology (S1)			
				Yes Pharmacy (S3)		No Pharmacy (S3)	
				Yes	No	Yes	No
General Practice (S4)	Yes	MNDA (S2)	Yes	6	0	6	1
			No	2	1	1	1
	No	MNDA (S2)	Yes	14	6	12	3
			No	21	7	2	?

estimates. The AIC (Akaike Information Criterion) is based on the goodness-of-fit statistic and can also help in model selection, the lowest AIC value indicating the best fitting model (1). The chosen model then provides the estimated number of cases not ascertained in the epidemiological survey. In this study a four-source capture-recapture analysis was performed and confidence intervals were calculated for the estimated number of missing cases (13). The adjusted incidence and prevalence rates are calculated simply by adding the missing cases to the actual number of identified cases and calculating as is performed for the crude rates.

Results

Incidence study

There were 67 incident cases of MND over the two-year study period: 35 males and 32 females. The neurology department ascertained 100% of cases while pharmacy ascertained 31% of cases, hospital coding 28%, EMG records 9%, GPs 4.5% and MNDA 6%. The average annual incidence rate of MND in NI was 2.0 per 100,000 person-years (95% CI 3.0–5.0) and 1.4 per 100,000 person-years (95% CI 0.9–2.0) when standardized to the European standard population (13). As all incident cases were captured by the neurology department and no cases were separately identified by any other source, completeness of ascertainment can be estimated at 100%.

Prevalence study

Eighty-three prevalent cases of MND were identified: 45 males and 38 females. Pharmacy ascertained 77% of cases while the neurology department ascertained 69% of cases, MNDA 59%, hospital coding 48%, EMG records 42%, and GPs 22%. The prevalence of MND in NI was 4.9 per 100,000 population (95% CI 3.8–5.9) and 3.3 per 100,000

Table II. Four-source log-linear models highlighting two-way interactions. All possible two-way interactions were examined using a statistical package (STATA) and their resultant G^2 and AIC values noted. The models containing the interactions S1/S2 and S1/S4 appeared to provide the best fit as assessed by the low G^2 and AIC values. Both models also compared favourably to the independence model that assumes no dependence between sources. A model containing both interactions was then performed and found to provide the best fit and therefore accepted as the model of choice. This model proposed that less than one case was missing from the prevalence survey. Importantly, however, the number of missing cases from any of the models was very small.

	G^2	p	AIC	Estimate of missed cases
Independence model	19.7	0.03	74.4	2.5
S1/S2	10.4	0.32	67.1	0.9
S1/S3	19.7	0.02	76.4	2.7
S1/S4	16.9	0.05	73.6	2.1
S2/S3	18.5	0.03	75.2	3.3
S2/S4	17.2	0.05	73.9	2.8
S3/S4	18.7	0.03	75.4	2.8
S1/S2, S1/S4	7.1	0.53	65.8	0.7

Chosen model in bold type.

G^2 : goodness-of-fit test; AIC: Akaike Information Criterion; S1: Neurology; S2: MNDA; S3: Pharmacy; S4: General Practice.

population (95% CI 2.4–4.2) when standardized to the European standard population (13). No cases were uniquely identified by EMG records or hospital coding lists and so these sources were excluded. The remaining four sources were then used for the analysis. These included the regional neurology department (S1), MND association (MNDA) (S2), the regional pharmacy department (S3), and GPs (S4).

Table I displays the data used for this analysis. All two-way interactions were examined for significance and the two-way interactions that attained significance included an interaction between neurology (S1) and the MNDA (S2), and an interaction between neurology (S1) and GPs (S4). The model including both interactions (S1/S2 and S1/S4) fitted the data better than models including only one or the other whether judged by the likelihood ratio statistic or the AIC. This model estimated that only 0.7 cases were missed. Although the principle of parsimony might favour the S1/S2 model, i.e. the simplest model that provides adequate fit, this model gave a very similar estimate of 0.9 missing cases (see Table II). The prevalence rate therefore remained virtually unchanged and the analysis suggested that the study had a coverage of 99% of all potential MND cases.

Discussion

This study employed capture-recapture analysis to estimate the number of missing cases for both the incidence and prevalence of MND in NI. The incidence

study showed that all cases were ascertained by one source and therefore provided an estimate of completeness of 100%. The prevalence study also revealed a high level of ascertainment, 99%, with a prevalence rate that remained virtually unchanged. Reassuringly, no matter which of the capture-recapture models was chosen, the number of cases estimated as missed was small. The study was felt to adequately satisfy the assumptions of capture-recapture analysis. All cases were 'true' cases, having been diagnosed by a consultant neurologist as well as being further examined by the author (CD) before entry into the study. All cases were accurately recorded, as were their source(s), by way of the NI MND register which simplified the matching process. The MND population can be considered to be relatively 'closed', as most patients are of retirement age with little likelihood of migration. The MND population is a reasonably homogeneous group and disease occurs almost exclusively in middle age in an almost equal number of both sexes. In addition, because of the severity of MND and the reasonably similar disease phenotype among patients at least at some point during their disease, patients tend to follow similar management paths. All patients should be seen by a neurologist and therefore be available for capture within the 'neurology source'. Patients diagnosed by neurologists in NI are put in touch with the MNDA, and this may account for the interaction and dependency observed in the analysis between these sources. The majority of patients should be available for capture within the 'pharmacy' source, as riluzole is the only licensed drug for MND. Finally, although the 'GPs' source has the potential to catch all patients with MND in NI it is not an efficient method of ascertainment. Nonetheless, the use of multiple different sources was likely to help overcome any disadvantage with respect to the catchability of cases. Due to the disabling nature of MND it is likely that in most countries, even where access to healthcare is variable such as in the privatized system in the United States, patients would present and be identified. However, in less progressive countries such as Africa, where access to health care is difficult, case ascertainment is challenging and problematic in general. Nonetheless, applying capture-recapture techniques to known incomplete sources provides the investigator with more knowledge than would otherwise be available and thus it is still useful. Where there is concern that socioeconomic status or other factors negatively impact upon ascertainment, patients can be stratified according to their socioeconomic status and separate analyses performed. This enables the investigator to establish if ascertainment varies between different subgroups. Finally, if concern remains surrounding 'equal catchability' of cases-within-sources, multiple sources can be pooled. This increases the chances of a true case being captured within the required source. To accommodate for any potential variation in geographical spread of

cases, the survey should be large enough to cover both rural and urban areas.

Few epidemiological surveys of MND have incorporated capture-recapture analysis to date (2-6). An incidence study from the Limousin region in France (2) used a three-source model employing a computerized ALS/MND neurology unit database, private practice neurologists, and data from hospitals serving the Limousin region. Estimating the number of missing cases increased the crude incidence rate from 3.2 to 4.9 per 100,000 population. However, because a large proportion of the Limousin region is over 65 years of age (22%); this latter figure reduced to 2.9 per 100,000 when standardized to the world population. The first prospective epidemiological study of MND in a Latin American country (6) reported good case ascertainment, with little resultant change made to their crude incidence rate. A study by Forbes et al. (4) used capture-recapture methodology to ensure that ascertainment bias had not affected their finding that the incidence of MND had not changed over a 10-year period. They found a high level of case ascertainment of between 96% and 99.3% throughout the 10-year period and showed that reduced ascertainment in more recent years was not responsible for the level secular trend in incidence. This was an interesting study because of numerous previous reports stating that MND incidence is increasing, and provides strong evidence that MND incidence is not increasing. Importantly, this finding was replicated in a more recent prospective Italian study of MND incidence and prevalence (3). Chio et al. (3) noted steady incidence rates over a 10-year period. They used a two-source capture-recapture analysis to confirm a high level of ascertainment with little resultant change made to crude disease rates. In contrast, a prospective epidemiological survey of MND in New Zealand (5), performed over a 22-year period, reported increasing incidence of MND. They also employed capture-recapture methodology; however, they were only able to apply this to the last 10 years. While they reported an increase in incidence of 1.6 to 3.3 per 100,000 based on a best-fit line drawn through their data, the observed data look less convincing, and some of the highest annual incidence rates appear to have been in the first 10 years of the study. In addition, because capture-recapture analysis was not available for the first 12 years of the study, it cannot be ruled out that poor case ascertainment played a part in lower incidence rates from the 1980s and early 1990s.

Overall, it seems reasonable that future investigators should employ this technique to help validate their estimations of disease frequency in MND and allow a better comparison to be made between different studies. Although employing capture-recapture methods may involve more planning, it is not necessarily more time intensive as it obviates the need to perform an exhaustive survey. Importantly, at a time where adherence to strict ethical policies on the processing of personal data can limit epidemiological

data collection, capture-recapture analysis may have a role to play. The presence of an MND register is likely to make the employment of capture-recapture easier. As the NI MND register continues, new measurements of disease frequency can be compared with the figures above. Any observed changes can then be evaluated using this methodology to ensure that case ascertainment does not account for apparent secular trends.

Acknowledgements

The authors express their appreciation to Alison Dick, MND Care Centre Coordinator and Noreen Kearney, MND Care Centre secretary, for their administrative assistance. This project was funded by a North-South project grant supplied by the R & D office.

Declaration of interest: The authors have no conflicts of interest.

References

1. Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev.* 1995;17: 243-64.
2. Chazot F, Vallat JM, Hugon J, Lubeau M, Dumas M. Amyotrophic lateral sclerosis in Limousin (Limoges area, France). *Neuroepidemiology.* 1986;5:39-46.
3. Chio A, Mora G, Calvo A, Mazzini L, Bottacchi E, Mutani R. Epidemiology of ALS in Italy: a 10-year prospective population based study. *Neurology.* 2009;72:725-31.
4. Forbes RB, Colville S, Parratt J, Swingler RJ. The incidence of motor neuron disease in Scotland. *J Neurol.* 2007;254: 866-9.
5. Murphy M, Quinn S, Young J, Parkin P, Taylor B. Increasing incidence of ALS in Canterbury, New Zealand: a 22-year study. *Neurology.* 2008;71:1889-95.
6. Vazquez MC, Ketzoian C, Legnani C, Rega I, Sanchez N, Perna A, et al. Incidence and prevalence of amyotrophic lateral sclerosis in Uruguay: a population based study. *Neuroepidemiology.* 2008;30:105-11.
7. Northern Ireland Statistics and Research Agency. www.nisra.gov.uk
8. The European Standard Population. www.nchod.nhs.uk
9. Capture-recapture and multiple-record systems estimation. I: History and theoretical development. International Working Group for Disease Monitoring and Forecasting. *Am J Epidemiol.* 1995;142:1047-58.
10. Le Cren E. A note on the history of mark-recapture population estimates. *J Anim Ecol.* 1965;34:453-4.
11. Wittes J, Sidel VW. A generalization of the simple capture-recapture model with applications to epidemiological research. *J Chronic Dis.* 1968;21:287-301.
12. Wittes JT, Colton T, Sidel VW. Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. *J Chronic Dis.* 1974;27:25-36.
13. Cormack RM. Interval estimation for mark-recapture studies of closed populations. *Biometrics.* 1992;48:567-76.