

The reliability of register-based patient characteristics

by

van den Akker M¹, Franssen G H L M², Buntinx F¹,
Metsemakers J F M¹, Knottnerus J A.¹

Abstract

To examine the reliability of demographical patient characteristics as routinely registered by the general practitioner (GP) in the Registration Network Family Practices, the educational level and type of living arrangement were compared with responders' answers in a postal questionnaire (N = 3 745). The degree of agreement was calculated using weighted and unweighted kappa values. The agreement for the level of education was good with an unweighted kappa value of 0.61 and a weighted kappa value of 0.75. The kappa value for type of living arrangement was also good with 0.73, but was considerably lower (0.43) for responders who did report any changes in living arrangement during the previous five years than for those who did not.

It is concluded that registration of patient characteristics by the GP was fairly reliable for the characteristics studied, but reliability gets worse in case of change. Therefore items that are subject to change, such as the type of living arrangement, should preferably not be used

Address Correspondence to: Marjan van den Akker, Department of General Practice, Maastricht University, P.O. Box 616, 6200 MD Maastricht.

¹ Maastricht University, Department of General Practice.

² Maastricht University, MEMIC, Centre for Data and Information Management.

for patient selection and should be measured again when important for a particular study.

Key-words

Bias, observer variation, registries.

Introduction

In the recent decades primary health care — as other professional areas in society — has developed in an atmosphere of advancing computer technology. At the same time an increasing number of general practice registrations were established, recording both health information and a more or less limited number of demographical patient characteristics. Such databases contain a lot of information and often provide possibilities for extensive analyses on e.g. the relationship between demographic features and morbidity or mortality, or are used as sampling frames for research.

Quality control of the registration systems varies. Metsmakers (1) described 13 sentinel stations and general practice registrations in the Netherlands. Most of these provide general feedback, but only half of them organise regular meetings of the registering general practitioners (GPs) to discuss (difficulties with) the registrations or guidelines, systematic errors or individual mistakes. Most studies on the reliability of registration focus on the recording of morbidity (2). But how well do GPs manage to provide reliable demographic data?

In this paper we report on an analysis comparing demographic information registered by general practitioners and self-reported demographics from a postal questionnaire.

Some characteristics are very stable and/or usually registered and updated by the GP, like age (date of birth) and type of health care insurance. Other characteristics might be relevant to the GP, but are too much subject to change to be registered reliably, e.g. employment. In this paper we concentrate on two characteristics that we expect to be either fairly stable at a certain age or known to the GP when changed and relevant to a sampling frame register: the level of education and the type of living arrangement.

Methods

Context

This study was carried out within the context of a large case-control study on the determinants of multimorbidity (N = 3 745): 1 368 cases were identified (subjects with two or more new diseases registered at the problem list in a three year period) and 2 377 controls were randomly selected (subjects with no or one new disease registered at the problem list in the same period). For this study the Registration Network Family Practices (RegistratieNet Huisartspraktijken = RNH) was used as a sampling frame. The RNH is a computerised and continuous database in which 42 Dutch GPs from 15 practices are participating. The GPs systematically collect all relevant chronic, recurrent or permanent health problems of over 60 000 patients. Additionally background information on sex, date of birth, living arrangement, level of education (low, middle, high) and type of health insurance (either private or public) of the patients is gathered. The registered data are continuously updated and historically cumulated for each patient. Population membership only ends by migration or death. All subjects included in the RNH are informed and have given their consent about possible use of their anonymized data. The quality of the data is ascertained by instruction and training sessions, regional consensus groups, quality control experiments and special software programs, such as an automated thesaurus and automated checking for erroneous or missing entries (3).

Subjects

Subjects that were selected for the case-control study received a postal questionnaire asking for living arrangements, weight and height, medical family history, education and occupation, lifestyle and a number of psychosocial measures. The questionnaire included two items which were also recorded in the RNH database: the level of education (low, secondary, high) and the type of living arrangement (living as a couple or in a family, single parent home, living alone, or other). The operationalisation of level of education and type of living arrangement was almost similar for GP and in the questionnaire. Only subjects could add free text for living arrangements which was later recoded to the existing categories. Only 2% of the responders used free text.

Analysis

All responders to the case-control study that answered the questions relevant for this study, and whose information was also available in the RNH database were included in the analyses.

The agreement between the RNH-recorded and self-reported features was analysed using Cohen's kappa (weighted and unweighted) with 95% confidence intervals (4).

For the level of education (three categories) unweighted kappa values and weighted kappa values using quadratic disagreement weights were calculated for all subjects, and separately for subjects under and over age 30.

For the type of living arrangement unweighted kappa values were calculated, for all subjects, and separately for subjects who reported recent change in living arrangement in the questionnaire (in the previous five years) and for those who did not report recent changes in living arrangement.

Results

Level of education

Information on the level of education was available for 3 451 subjects (92.1%). In 2 647 cases the GP and the respondent agreed on the level of education (Table 1). For 286 subjects the GP recorded a higher level of education than the respondent did, whereas in 518 cases the GP recorded a lower level of education than the respondent.

TABLE 1
RNH-recorded and self-reported level of education, absolute numbers (N = 3 451)

		Questionnaire			
		Low	Secondary	High	Total
Register	Low	1 440	320	16	1 776
	Secondary	216	881	182	1 279
	High	3	67	326	396
	Total	1 659	1 268	524	3 451

The unweighted kappa value was 0.61; the weighted kappa value was 0.75. The weighted kappa value for subjects of 30 years or older hardly differed, but was lower (0.65) for subjects younger than 30 (Table 2).

TABLE 2
Unweighted and quadratic weighted kappa values for level of education

Subjects	N	Unweighted kappa (95% CI)	Weighted kappa (95% CI)
All subjects	3 451	0.61 (0.60-0.62)	0.75 (0.73-0.77)
Subjects younger than 30	260	0.51 (0.46-0.57)	0.65 (0.61-0.69)
Subjects aged 30 and older	3 191	0.61 (0.60-0.63)	0.75 (0.74-0.77)

Living arrangement

Data regarding the type of living arrangement were available for 3 615 subjects (96.5%). Table 3 shows that for 88.7% (N = 3 208) of the subjects similar information has been provided by responders and GPs with respect to the responders' living arrangement.

TABLE 3
*RNH-recorded and self-reported type of living arrangement, absolute numbers
(N = 3 615)*

		Questionnaire				Total
		Family/ couple	Single parent	Alone	Other	
Registered	Family/couple	2 715	31	141	11	2 898
	Single parent	33	31	39	6	109
	Alone	101	8	437	21	567
	Other	3	0	13	25	41
	Total	2 852	70	630	63	3 615

The unweighted kappa value was 0.73. When the kappa values were calculated separately for subjects who reported changes in living arrangement during the previous five years and those who did not, kappa values were 0.43 and 0.83 respectively (Table 4).

TABLE 4
Kappa values for type of living arrangements

Subjects	N	Kappa	95% confidence interval
All subjects	3 615	0.73	0.71-0.76
No recent change ¹	2 982	0.83	0.81-0.86
Recent change ¹	633	0.43	0.35-0.50

¹ Change in living arrangement during the last five years as reported in the questionnaire.

Discussion and conclusion

According to Fleiss (5) a kappa value higher than 0.75 reflects excellent agreement, whereas kappa values between 0.40 and 0.75 indicate fair to good agreement above chance. Kappa values lower than 0.40 are classified as poor.

This means that in this study the agreement for the level of education was good. As expected the agreement was lower amongst younger people (under age 30). The educational level of younger people more often changes because they still receive education, which increases the chance that the GP's information is not up to date.

Also the agreement regarding the type of living arrangement was good and even excellent for subjects who reported no changes in living arrangement during the last five years. For responders reporting recent changes in living arrangement however, agreement was only just fair.

This indicates that the GP has good notion of the educational level of his patients, but has trouble keeping abreast of changes of living arrangement, like marriage, cohabiting and divorce in his files. The GP is possibly not always informed on changes or does not consistently register this type of demographic changes.

When analysing data or selecting subjects from a registration system using demographic characteristics as selection criteria, this should preferably be limited to features like age and sex which are not affected by alterations in time. When a registration system is used as a sampling frame for patient based research we suggest that the relevant patient characteristics that are amenable to change are measured again by interview or questionnaire. Researchers should at

least be aware of the level of reliability of patient characteristics when using them.

A limitation of this study is the fact that double information was only available for two patient characteristics. Further reports into the reliability of routinely registered patient characteristics are useful, because results might have implications for primary care based registrations as well as for registration based studies.

Acknowledgements

This study was supported by the Dutch Praeventiefonds (28-2692). The authors thank HCW de Vet from the Department of Epidemiology of Maastricht University for her help regarding software and analysis of the weighted and unweighted kappa values.

Samenvatting

Doel van deze studie was de betrouwbaarheid vast te stellen van de routinematige registratie van achtergrondkenmerken van patiënten door hun huisarts. Hiertoe zijn het opleidingsnivo en het type woonverband zoals geregistreerd in het RegistratieNet Huisartspraktijken (RNH) vergeleken met gegevens die patiënten hierover zelf hebben gegeven in een schriftelijke vragenlijst (N = 3 745). Om de mate van overeenstemming te bepalen zijn gewogen en ongewogen kappa-waarden berekend. De overeenstemming ten aanzien van het opleidingsnivo was goed, met een ongewogen kappa-waarde van 0,61 en een gewogen kappa-waarde van 0,75. De kappa-waarde voor het type woonverband was ook goed met een waarde van 0,73, maar deze was aanzienlijk lager (0,43) voor patiënten die in de postenquête aangaven dat hun woonverband veranderd was gedurende de laatste 5 jaren.

Uit de resultaten blijkt dat de registratie van patiëntkenmerken door de huisarts tamelijk betrouwbaar is, maar dat de betrouwbaarheid minder wordt bij mutaties. Bij voorkeur zouden mutatiegevoelige kenmerken, zoals woonverband, daarom niet gebruikt moeten worden voor de selectie van patiëntengroepen. Deze kenmerken zouden bovendien opnieuw gemeten dienen te worden als ze belangrijke informatie vormen voor een bepaald onderzoek.

References

1. METSEMAKERS J F M. Unlocking the patients' records in general practice for research, medical education and quality assurance: the Registration Network Family Practices. Amsterdam: Thesis Publishers Amsterdam 1994.

2. BARTELDI A I. Validation of sentinel data. *Gesundheitswesen* 1993; 55 (1 Suppl.): 3-7.
3. METSEMAKERS J F M, HÖPPENER P, KNOTTNERUS J A, KOCKEN R J J, LIMONARD C B G. Computerized health information in the Netherlands: a registration network of family practices. *Br J Gen Practice* 1992; 42: 102-106.
4. SCHOUTEN H J A. Nominal scale agreement among observers. *Psychometrika* 1986; 51: 453-466.
5. FLEISS J L. *Statistical methods for rates and proportions*. New York: Wiley 1981.