Validity of self-reported hospital admissions in clinical trials depends on recall period length and individual characteristics

Authors

Hildegard Seidl[1], Christa Meisinger[1, 2], Inge Kirchberger[1,2] , Katrin Burkhardt[2], Bernhard Kuch[3], Rolf Holle[1]

[1] Helmholtz Zentrum München - Institute of Health Economics and Health Care Management, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany

[2] Augsburg Hospital, MONICA/KORA Myocardial Infarction Registry, Stenglinstraße 2, 86150 Augsburg, Germany

[3] Department of Internal Medicine/Cardiology, Donau-Ries-Kliniken, Stoffelsberg 4, 86720 Nördlingen, Germany

Contact of the corresponding author:

Hildegard Seidl, Economist (BSc), MPH

Research associate

Helmholtz Zentrum München
Institute of Health Economics and Health Care Management
Ingolstädter Landstraße 1
D-85764 Neuherberg

Germany
Tel.: +49 (0)89/3187-4485
Fax: +49 (0)89/3187-3375

email: hildegard.seidl@helmholtz-muenchen.de

Abstract

Rationale, aims and objectives

We investigated the validity of self-reported admission data compared to administrative records in a clinical trial.

Method

In the randomised KORINNA study (ISRCTN02893746), hospital admission data were collected in telephone interviews with 273 elderly patients quarterly over a 1-year period and thereafter annually over a 2-year period. Data were compared with administrative records and discharge letters. Mixed models were used to investigate if recall period and individual characteristics influence validity.

Results

Specificity (>99%) and sensitivity (94%) of self-reported data did not differ for different recall periods (3 months vs. 12 months). The differences between self-reported and registered inpatient days were not statistically significant. Having regard to all the admissions within the time period of last interview and dropping out, the bias was up to 40% underestimation. The chance of disagreement was significantly smaller (OR of misremember an admission=0.596, p=0.049, CI=0.355 to 1.00; OR of misremember length of stay=0.521, p=0.002, CI=0.344 to 0.789) for 3-month periods; but this was primarily driven by number of admissions within the recall period. Individuals with better health and longer stays had a significantly smaller chance of disagreement.

Conclusions

The bias within 1 year was not influenced by applying various recall periods although the probability of correctly self-reported single hospital admission was higher using a recall period of 3 months. It can be recommended that lengthened recall periods of 12 months are appropriate for gathering self-reported hospital admission data in elderly people with myocardial infarction.

Keywords: Validity**,** Self-reported hospital admissions**,** Recall period**,** Recall bias**,** Elderly

Manuscript

Validity of self-reported hospital admissions in clinical trials depends on recall period length and individual characteristics

1 Introduction

Economic evaluation studies are based on healthcare costs. Especially in randomised controlled trials (RCT), healthcare costs often rely on self-reported data (1, 2). Due to patients' enrollment in different insurance funds, cost data collection from health insurance companies is complex or may not be possible. Moreover, self-reported resource use offers an account of out-of-pocket payments and informal health services, which are relevant for economic evaluation from a societal perspective. Therefore, the question arises as to how valid is self-reported resource use. Missing data and recall error affect the validity of self-reported resource use. Recall error is also influenced by length of recall period, frequency of use, or patient-related aspects such as age, gender, or health state and cognitive ability (3). The validity of self-reported hospital admissions is of particular importance as hospital admissions are often the main cost drivers. The demographic trend towards an ageing population, which is a relevant driver for hospital care (4), calls for further research on the validity of self-reports in elderly patients.

The question on the impact of recall period on the extent of recall bias has specific relevance in RCTs with longer follow-up periods. In studies with a follow-up period of 1 year, for example, participants can be either asked annually or more frequently with shorter recall periods, increasing ,however, the burden on participants (5). Prolonging the time between participant contacts could increase missing values resulting from the increased chance of dropping out before being contacted. Reduced data collection from shortened recall periods and then extrapolating to the target period may not be a suitable option because information are less precise than asking for the whole target period (6, 7). Therefore, more empirical evidence is needed to enable health economic researchers to control the impact of recall errors and recall bias when choosing the most appropriate design for data collection in RCTs or other longitudinal studies.

A few studies have validated self-reported healthcare use in the elderly population for a range of healthcare variables, but the percentage of participants who had at least one hospital admission was often only 3 – 20% (8-12). Therefore, results on the accuracy of self-reported hospital admissions were driven primarily by non-users. Moreover, participants of population based studies usually are healthier than participants of RCTs so that generalisability from general population to RCT participants is limited. There are two longitudinal studies in patients with mental disorders but without considering different recall periods (13, 14) and one study in primiparous women but disregarding hospital admissions in the regression analysis because of insufficient prevalence of admissions (15). Only one population based Swedish study (6) examined the impact of different recall periods on the validity of self-reported hospital admissions, and one population based US study (16) reported agreement separated into monthly and annual recall periods.

The aim of this study is to investigate recall error and recall bias with respect to self-reported hospital admissions in a typical RCT in elderly patients with myocardial infarction and to analyse how length of recall period and individual characteristics such as gender, age, cognitive ability, and health status influence the validity of each self-reported single admission.

2 Methods

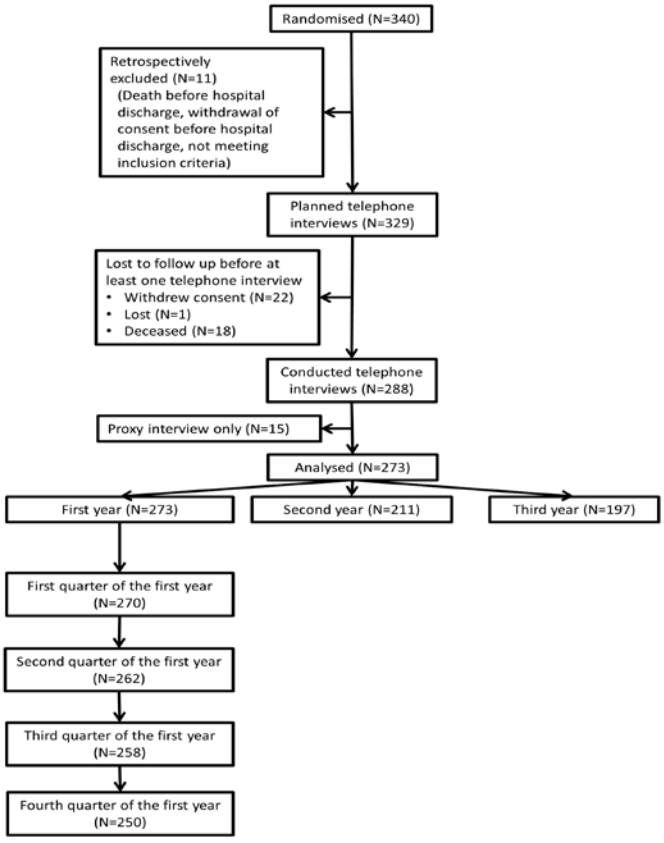
We distinguish between recall error and recall bias. Recall error occurs as a result of incorrect self-report, i.e. over- or underreport of a single admission or length of stay. In contrast, when several incorrectly reported admissions or length of stays do exist, recall bias only occurs if overreporting outweighs underreporting (or vice versa) otherwise errors may cancel each other out.

2.1 Sample/study population

The data were obtained from the randomised controlled KORINNA trial, which evaluated the effectiveness and cost-effectiveness of a case management intervention by trained nurses in elderly patients with an acute myocardial infarction. The study design, intervention and results of the KORINNA trial are described elsewhere (17-20). The trial was approved by the Ethics Committee of the Bavarian Chamber of Physicians. The trial registration number is ISRCTN02893746. In short, between September 2008 and May 2010, 340 patients from the Augsburg region were recruited who

were at least 65 years old and had an acute first or recurrent myocardial infarction treated at the Central Hospital of Augsburg (CHA). This is the major hospital – a tertiary care unit with 1731 beds – for the population of 830,000 in the Greater Augsburg area, Germany. Exclusion criteria were a planned or present residence in a nursing home, severe comorbidity associated with a life expectancy of less than 1 year, insufficient ability to speak German and lack of ability due to cognitive disorders or willingness to consent. Patients who died or withdrew consent before hospital discharge were retrospectively excluded (n=11). The intervention consisted of at least one home visit and quarterly telephone calls during the first year and semiannual calls during the second and third years. The intervention and observation period spanned 3 years. For investigating validity of self-reported hospital admissions we excluded 15 participants because of conducting only proxy interviews; moreover, 41 participants had no interview because of dropping out or early death. Therefore a total of 273 participants attended at least one telephone interview.

Figure 1 gives an overview of conducted telephone interviews in the respective follow-up period.



2.2 Data collection

In the first year, quarterly telephone interviews were conducted to collect data about healthcare use including hospital admissions in the past 3 months. In the second and third years, annual interviews were conducted covering the past 12 months. We asked for admission (no/yes) and, if yes, we asked for number of admissions; and for each admission, we asked for days per admission (length of stay), diagnosis, admission date, name and location of the hospital, and whether it was an unplanned or planned admission.

Each self-reported admission was validated by the study physician regarding date of admission, diagnosis, length of stay, and whether hospitalisation was planned or unplanned using hospital records or discharge letters. For the CHA, all hospital records were available and extracted for the whole study period and for every participant. For every participant who reported at least one admission to any other hospital all hospital records and discharge letters were requested from this hospital for the whole study period. If participants were transferred from one hospital to another, then this was counted as a single admission. Self-reported length of stay was regarded as valid if reported days corresponded with registered days of the respective admission. To quantify the problem of uncovered hospital admissions based on participants dropping out before being contacted we consider all admissions within the period of time between last interview and dropping out. Thus our data set consisted of each reported or unreported but registered single admission.

Functional status and cognitive ability were assessed at baseline and 1 year after discharge; depressive symptoms were also assessed 3 years after discharge, and self-rated health state was assessed at every interview.

Cognitive ability was measured using the Mini-Mental State Examination (MMSE) (21). Scores lower than 24 are considered to be indicative of cognitive impairment (22) and dichotomised accordingly.

Depressive symptoms were assessed using the 15-item version of the Geriatric Depression Scale (GDS-15) (23). Scores above 5 are used to indicate clinically important depressive symptoms (24) and dichotomised accordingly.

Patients' self-rated health state was assessed using the EuroQol five-dimensional questionnaire visual analogue scale (EQ-VAS) ranging from 0 (worst imaginable health state) to 100 (best imaginable health state) (25).

Functional status was examined using the Barthel Index (26) ranging from 0 (totally dependent) to 100 (totally independent).

The Barthel-Index and MMSE were administered through medical examination and scored by the study physician; GDS-15 and EQ-VAS were administered through personal interviews conducted by the study nurse.

2.3 Statistical analysis

Descriptive statistics are presented for all patient characteristics. The prevalence of registered admissions and self-reported admissions is reported for each interview time point and recall period.

Information from discharge letters and hospital records was regarded as the gold standard for the calculation of the sensitivity, specificity, and accuracy of reports in the respective time period. Sensitivity was calculated as the percentage of participants who accurately reported at least one hospital admission among all participants with at least one registered hospital admission. Specificity was calculated as the percentage of participants who accurately reported no hospital admission among all participants without registered hospital admissions. Accuracy was calculated as the percentage of correctly reported admission information among all patients.

The mean difference (bias) between self-reported and registered hospital days summed up within 1 year was analysed using a paired t-test. Alternatively, bootstrap resampling with 1000 resamples was applied to compute p-values because of the skewed distribution of the difference. The extent of under- and overreported hospital days within 1 year are presented as means; the absolute error and bias are presented as means plus relative values based on registered hospital days. We performed two separate calculations of bias, one due to participants' reporting error only, and one including admissions between the last interview and dropping out.

To investigate the influence of different recall periods and individual characteristics on recall error, logistic mixed models with subject as a random effect and compound symmetry dependence were

fitted. Mixed models have been shown to be an effective method in longitudinal studies with repeated measurements since they can deal with correlated within-subject errors (27, 28). In our study, participants were contacted four times in the first year and once a year in the second and third years.

We fitted three multivariable logistic mixed models (model 1) and calculated odds ratios (OR) to capture three different recall errors: disagreement in the single admission, disagreement in days per stay, and underreporting an admission, with the following classifications: 'underreporting an admission' indicates that a registered admission was not reported; 'disagreement in the single admission' indicates an underreport or that a reported admission was not registered; and 'disagreement in days per stay' indicates that self-reported length of stay did not correspond with registered length of stay including disagreement in admission. The independent variables were recall period (quarterly vs. annually), GDS (depressive symptoms vs. no depressive symptoms), MMSE (cognitive impairment vs. no cognitive impairment), age at baseline (<70 vs. 70–79 vs >80), gender, Barthel Index score (continuous 1–100), EQ-VAS (continuous 1–100). Depressive symptoms, cognitive impairment, and health state were included since they have shown to influence recall ability (3, 6, 12, 13, 15). The Barthel Index measures the functional status and reflects the objective health state whereas the EQ-VAS reflects the subjective self-rated health state. Since the independent variables were measured several times, the respective values closer to the time of the conducted interview were used in the regression. Furthermore, we included length of stay because we assumed that the longer the stay, the more salient the admission and that participants might be able to remember better (3). In an additional analysis (model 2), we included number of admissions during the recall period because the more admissions the participants have, the greater the likelihood of misreporting (3) could be. All analyses were performed using SAS (Version 9.2, SAS–Institute Inc., Cary, NC, USA).

3 Results

3.1 Sample characteristics

The mean age was 75.0 years and 35.5% were female. Self-rated health according to EQ-VAS was on average 64.4 on a rating scale from 0 to 100; 16.5% had clinically important depressive symptoms and 12.8% were cognitively impaired (Table 1).

Table 1 - Characteristics of study participants at baseline, (n=273)

Age (mean [SD])	75.0	[5.7]
Sex (% females)	35.5	
Barthel Index (mean [SD])	91.8	[16.2]
EQ-VAS (mean [SD])	64.4	[19.2]
Presence of depressive symptoms (%)	16.5	
Presence of cognitive impairment (%)	12.8	

3.2 Agreement

Table 2 presents the prevalence of admission and its agreement. In the first year after suffering from an acute myocardial infarction, about half the participants who attended at least one interview had at least one hospital admission, which decreased to 37% in the second year and 35% in the third year. Specificity of self-reported admissions was generally above 99%, independent of length of recall period. Sensitivity was 93.85% for all four recall periods of 3 months combined in the first year, and 91.18% and 96.20% for the second and third years (recall period of 12 months). Accuracy was above 96% for all recall periods.

Table 2 - Agreement in prevalence of registered and self-reported admissions

year and length of recall period	registered and reported	registered but not reported	not registered and not reported	not registered but reported	sensitivity	specificity	accuracy
	numbers [%]	numbers [%]	numbers [%]	numbers [%]	%	%	%
1st year, 1st quarter (n=270)							
3 months	52 [19.26]	4 [1.48]	212 [78.52]	2 [0.74]	92.86	99.07	97.78
1st year, 2nd quarter (n=262)							
3 months	55 [20.99]	3 [1.15]	198 [75.57]	6 [2.29]	94.83	97.06	96.57
1st year, 3rd quarter (n=258)							
months	31 [12.02]	5 [1.93]	220 [85.27]	2 [0.78]	86.11	99.10	97.29
1st year, 4th quarter (n=250)							
months	33 [13.2]	5 [2.0]	211 [84.4]	1 [0.4]	86.84	99.53	97.60
first year							
sum of 4×3months (n=273)	122 [44.69]	8 [2.93]	142 [52.01]	1 [0.37]	93.85	99.30	96.70
second year							
12 months (n=211)	76 [36.02]	3 [1.42]	131 [62.09]	1 [0.47]	96.20	99.24	98.10
third year							
12 months (n=197)	62 [31.47]	6 [3.04]	128 [64.98]	1 [0.51]	91.18	99.22	96.45

Note. Registered and reported means that both the hospital records and the participant displayed at least one admission; registered but not reported means that the hospital records displayed at least one admission but the participant did not; not registered and not reported means that neither the hospital records nor the participant displayed an admission; not registered but reported means that the hospital records did not display an admission but the participant did.

3.3 Extent of bias

Table 3 shows differences in days within one year and the resulting bias. On average, participants reported 8.28 hospital days compared with 7.87 registered hospital days in the first year. In the second and third years, 4.39 and 4.31 hospital days, were reported compared with 4.58 and 4.25 registered days. The bias was not statistically significant and the relative bias was 5% or less over both recall periods.

The mean absolute error is higher if applying recall periods of 3 months (summing up the four 3-month recall periods in the first year) compared with the 12-month recall period from years 2 and 3 (1.79 vs. 0.64 and 0.66). This is also true for the relative absolute error (23% vs. 14% and 16%). The higher relative absolute error in the first year is more driven by overreporting (14%) than underreporting (9%).

If admissions between the last interview and dropping out were included, the bias was statistically significant. It led to an underestimation of 2.27 days (summing up the four 3-month recall periods in the first year), 2.4 days (interviewing annually in the second year), and 1.29 days (interviewing annually in the third year). This equated to a relative bias of 24% in the first year, 38% in the second year, and 24% in the third year. Translated into estimated mean costs (€593.03 per hospital day (29)), the underestimation was €1,346 in the first year, €1,423 in the second year, and €763 in the third year.

I

Table 3 - Comparison and differences in self-reported and registered days without drop-outs and including drop-outs

	length of recall period	**self-reported** mean, [SD]	**registered** mean, [SD]	**differences in days**				
				underreported days mean	**overreported days** mean	**absolute error** mean, (relative error in %, based on registered days)	**bias** mean, [SD], (relative bias in %, based on registered days)	p-values of bias (bootstrapping)
without drop-outs								
sum of first year (n=273)	3 months	8.28 [17.14]	7.87 [14.52]	–0.69	1.10	1.79 (23)	0.40 [0.49] (5)	0.3394 (0.3450)
second year (n=211)	12 months	4.39 [9.93]	4.58 [10.03]	–0.41	0.23	0.64 (14)	–0.19 [2.78] (4)	0.3293 (0.3810)
third year (n=197)	12 months	4.31 [9.62]	4.25 [9.43]	–0.31	0.36	0.66 (16)	0.05 [2.81] (1)	0.7901 (0.7790)
including drop-outs								
sum of first year (n=314)	3 months	7.30 [16.33]	9.57 [17.14]	–3.22	0.95	4.17 (44)	–2.27 [11.04] (24)	0.0003 (<0.0001)
second year (n=237)	12 months	3.91 [8.53]	6.31 [13.17]	–2.60	0.20	2.80 (44)	–2.40 [10.11] (38)	0.0003 (0.0150)
third year (n=203)	12 months	4.08 [9.40]	5.36 [11.71]	–1.63	0.34	1.96 (37)	–1.29 [8.25] (24)	0.0256 (0.0720)

Note. Underreported days = reported days – registered days if reported days < registered days, otherwise underreported days = 0;
overreported days = reported days – registered days if reported days > registered days, otherwise overreported days = 0;
absolute error = absolute difference between reported and registered days
bias = reported days – registered days

3.4 Influence on recall error

The results of multivariable logistic mixed model regressions are shown in table 4. Without controlling for number of admissions (model 1), the chance of disagreement in the single admission was significantly smaller (OR=0.596, p=0.0498, CI=0.355; 1.00) for individuals being asked for the previous 3 months compared with those being asked for the previous 12 months and for individuals with depressive symptoms compared with those without depressive symptoms (OR=0.431, p=0.0493, CI=0.186; 0.997). For a person with a 1-point higher EQ-VAS score (i. e. better health), the chance of disagreeing was 0.97 significantly smaller (p=0.0002, CI=0.955; 0.986). The variables age, gender, MMSE, Barthel Index score, and length of stay in days did not significantly explain disagreement in admission. Disagreement in days was significantly explained by recall period (OR=0.521, p=0.0021, CI=0.344; 0.789), EQ-VAS score (OR=0.979, p=0.0005, CI=0.967; 0.991), and length of stay (OR=1.061, p=<0.0001, CI=1.035; 1.088).

The chance of underreporting an admission was also significantly associated with recall period (OR=0.516, p=0.0471, CI=0.268; 0.991) and EQ-VAS score (OR=0.966, p=0.0007, CI=0.947; 0.985). For a person who was asked to recall the previous 3 months, the chance of underreporting was 0.516 smaller than asking this person for the previous 12 months.

When controlling for number of admissions during the recall period (model 2), results of disagreement and underreporting changed. The chance of disagreement in admission (OR=2.047, p=<0.0001, CI=1.593; 2.630) and in days (OR=1.768, p=<0.0001, CI=1.469; 2.127) increased with one additional registered admission during the recall period – driven by increased chance of underreporting (OR=2.601, p=<0.0001, CI=1.833; 3.693) – while the coefficient for recall period changed and became insignificant. The estimators of EQ-VAS, MMSE, and age remained robust. The estimators' significance of gender, depressive symptoms, and length of stay partially changed in the respective models.

Table – 4 Influence of variables on disagreement in the single admission and in days per stay, and on underreporting, OR of logistic mixed model regressions (p-values) [CI]

	disagreement (n=1603)[a]				underreporting an admission (n=1603)[a]	
	in the single admission		in days per stay			
	model 1	model 2	model 1	model 2	model 1	model 2
Recall period (ref=annual)	0.596 (0.0498)* [0.355; 1.00]	1.092 (0.7726) [0.600; 1.989]	0.521 (0.0021)* [0.344; 0.789]	0.875 (0.5720) [0.552; 1.389]	0.516 (0.0471)* [0.268; 0.991]	1.184 (0.6773) [0.543; 2.623]
Age 70–79 years (ref=<70 years)	1.389 (0.5028) [0.531; 3.633]	1.332 (0.5383) [0.534; 3.319]	1.089 (0.8106) [0.543; 2.182]	1.060 (0.8653) [0.542; 2.072]	0.813 (0.7204) [0.262; 2.526]	0.718 (0.5611) [0.234; 2.198]
Age >80 years (ref=<70 years)	1.739 (0.3349) [0.564; 5.361]	1.571 (0.4105) [0.536; 4.605]	1.767 (0.1696) [0.784; 3.984]	1.638 (0.2650) [0.749; 3.584]	0.973 (0.9680) [0.250; 3.783]	0.721 (0.6408) [0.183; 2.849]
Gender (ref=male)	0.608 (0.1687) [0.299; 1.234]	0.519 (0.0583) [0.263; 1.023]	0.659 (0.1201) [0.389; 1.115]	0.595 (0.0458)* [0.357; 0.990]	0.419 (0.0718) [0.163; 1.080]	0.283 (0.0136)* [0.104; 0.774]
GDS (ref=no depressive symptoms)	0.431 (0.0493)* [0.186; 0.997]	0.513 (0.1163) [0.223; 1.180]	0.603 (0.1155) [0.321; 1.132]	0.688 (0.2364) [0.370; 1.278]	0.446 (0.1359) [0.155; 1.289]	0.556 (0.2912) [0.187; 1.654]
MMSE (ref=no cognitive impairment)	2.001 (0.1293) [0.816; 4.903]	1.881 (0.1535) [0.790; 4.479]	1.486 (0.2892) [0.714; 3.093]	1.499 (0.2650) [0.735; 3.056]	2.394 (0.1215) [0.793; 7.227]	2.721 (0.0763) [0.900; 8.228]
Barthel Index	0.991 (0.3002) [0.973; 1.008]	0.989 (0.2332) [0.972; 1.007]	0.991 (0.2188) [0.977; 1.005]	0.992 (0.2287) [0.978; 1.005]	0.993 (0.5595) [0.971; 1.016]	0.991 (0.4556) [0.969; 1.014]
VAS of self-rated health state	0.970 (0.0002)* [0.955; 0.986]	0.978 (0.0072)* [0.963; 0.994]	0.979 (0.0005)* [0.967; 0.991]	0.986 (0.0216)* [0.974; 0.998]	0.966 (0.0007)* [0.947; 0.985]	0.975 (0.0166)* [0.955; 0.995]
Length of stay in days (registered)	0.966 (0.1427) [0.921; 1.012]	0.907 (0.0039)* [0.850; 0.969]	1.061 (<0.0001)* [1.035; 1.088]	1.034 (0.0100)* [1.008; 1.061]	1.013 (0.5546) [0.971; 1.056]	0.962 (0.1926) [0.907; 1.020]
Number of registered admissions during the recall period		2.047 (<0.0001)* [1.593; 2.630]		1.768 (<0.0001)* [1.469; 2.127]		2.601 (<0.0001)* [1.833; 3.693]

[a]n=1603, a repeated measures, analysis on 273 participants
Model 1: without number of registered admissions during the recall period
Model 2: including number of registered admissions during the recall period
OR= odds ratio; CI = Confidence Intervals; GDS = Geriatric Depression Scale; MMSE=Mini-Mental State Examination; VAS=Visual Analogue Scale
*=p-value -<0.05

4 Discussion

This study examined recall error and recall bias by comparing self-reported admission data with administrative records and discharge letters in the KORINNA study. We quantified the problem of uncovered hospital admissions due to participants dropping out before being contacted. In addition, we investigated if validity of self-reported hospital admission depends on recall periods and patient characteristics in elderly people with myocardial infarction attending a clinical trial.

We found high sensitivity, specificity, and accuracy in both a recall period of 3 months and a recall period of 12 months. The relative bias, caused by incorrect self-reports, (Table 3) in the 3-month recall periods (5%) was insignificantly higher than the bias in the 12-month recall periods (4% and 1%) whereas the relative bias, caused by participants dropping out, amounted to 24% and 38% in the 3-month recall period and 12-month recall period, respectively. The chance to misremember an admission or days per stay was significantly smaller for 3-month periods as well as for individuals with better health and longer stays but also for individuals with depressive symptoms compared with those without depressive symptoms.

Strengths of the study are firstly, that the RCT included a relevant population in which hospital admissions are the main cost driver. Thus, it is crucial to assess recall bias in the respective population rather than in the general elderly population which is not representative of elderly patients with myocardial infarction. Secondly, because of the sample size and the balanced number of users and non-users, the results were not driven primarily by non-users. Thirdly, the existing information about participants, such as health and functional status, allowed for investigating the influence of individual characteristics on validity. Fourthly, the detailed information about each single admission enabled matching each reported admission with registered admission. Other researchers (6, 13) only had information about numbers of admissions or the sum of days of all admissions during the recall period but no information about the length of stay per admission. Fifthly, by the presence of two different lengths of recall period we are able to investigate if recall bias depends on recall period length.

A limitation of our study is that allocation to different length of recall periods was not at random in the KORINNA study. All participants were first exposed to recall periods of 3 months over a year and then to recall periods of one year. This may introduce two issues which affected the internal validity. Firstly, it may be postulated that repeated reporting leads to a learning effect. In this case we would expect that the recall bias is associated with the different quarters in the first year. To test this, we used a logistic mixed model to examine if there is a different influence of the four 3-months recall periods on disagreement in the first year. The estimators of the four recall periods did not show any trends indicating that there was no learning effect. Secondly, more admissions occurred in the first year than in the second and third years (Table 3). Therefore, we do not know how valid self-reported admissions would have been when asking annually in the first year, especially since there was an association between misreporting and number of admissions. In order to gain more insight we compared the prevalence of admissions and mean number of admissions over the different years. While prevalence of admission differed significantly between first year (47%) and second (37%) and third (35%) years, mean number of admissions for patients having at least one admission did not differ significantly (1.84 vs. 1.65 vs. 1.59). Therefore, it can be assumed that results would remain robust when participants have been asked annually in the first year.

Two further limitations affect the external validity. Firstly, the participants of our study had an acute cardiovascular event so that our sample is not representative of the general elderly population but of the elderly population with heart disease. Secondly, considering that participants were asked to report detailed information about their hospitalisation (such as number of admissions, and for each admission length of stay, etc.) and that these questions could act as memory cues, our results would only be partially comparable to other studies. We cannot rule out the possibility of a memory cue given the high sensitivity and specificity despite the participants' sickness. It is worthwhile to consider including such additional questions to improve validity of self-reported hospital admissions in other studies.

Sensitivity and specificity are comparable to other studies which included elderly people. Wolinsky et al. (12) found a similar sensitivity of 92.66 but a lower specificity of 92.93. In contrast, Wallihan et al. (30) reported both a lower sensitivity of 75.25 and a specificity of 92.19. Raina et al. (10) also reported a lower sensitivity (80.46) but a similar specificity (96.29). Similar to our results, two studies which

included only elderly people (10, 12) have observed that individuals with better health states reported their admission more accurately. In contrast, Yu et al. (31), who included all age groups, found a correlation between poor health and the likelihood of accuracy. It is noticeable that we found higher validity in healthier individuals, but we also found higher validity in participants with depressive symptoms. Rozario et al. (14) did not find a significant association between depressive symptoms and the accuracy of self-reported hospital admissions, whereas Wolinsky et al. (12) reported a higher chance of underreporting in depressed participants. Our findings that men were more likely to misreport their hospital admissions are consistent with Wolinsky et al. (12). Other researchers found no gender association (6, 13, 14, 16).

We considered hospital records and discharge letters as the gold standard for calculating sensitivity and specificity, because reimbursement is based on these records and data quality can be assumed to be high. A restriction is that hospital records and discharge letters were not available for all hospital sites. In the case of admission to CHA, all hospital records were available so that all unreported hospital admissions could be detected. This cannot be assumed for other hospitals as a participant must be admitted at least once to that hospital to detect an unreported admission. About 9% of the admissions to CHA were not reported compared with 4% of non-reported admissions to other hospitals. As it is expected that the percentage should be equal, up to eight admissions may have remained undetected throughout the 3 years. Incorporating these admissions, the sensitivity would slightly decrease to 91.05 (first year), 93.8 (second year), and 88.6 (third year) and the specificity to 99.28 (first year), 99.23 (second year), and 99.21 (third year).

A very important aspect of our findings is the bias caused in longitudinal studies by participants dropping out before being (re-)contacted. We assumed that collecting hospital admission data quarterly would ensure a small bias due to participants dropping out. However, the relative bias amounted to 24% and caused a relevant underestimation albeit applying short recall periods. Applying a 12-month recall period in the second year, the relative bias increased to 38% although the drop-out rate was almost the same (24% in the first year, 23% in the second year).

The estimator for recall period changed and became non-significant when controlling for number of admissions during the recall period since disagreement was primarily driven by the number of

admissions. The longer the period the more admissions are naturally possible so that in this way the length of recall period influences recall error rather than length of period itself. Kjellsson et al. (6) reported a significantly reduced chance ((exp(–0.030)=OR of 0.970) of disagreement in prevalence of admission when asking for the previous 3 months instead of 12 months, although adjusting for registered admission. However, disagreement in prevalence of admission is not similar to disagreement in the single admission and therefore hardly comparable. We found that disagreement in the single admission was significantly lower for longer stays (model 2, OR=0.907). This could be assumed as longer stays represent salient events which are more likely to be reported accurately (3).

Further research should be done to investigate the influence of recall periods on validity of self-reported hospital admission in clinical trials by randomised allocation to different length of recall periods.

5 Conclusions

The bias within 1 year was not influenced by applying various recall periods although the probability of correctly self-reported single hospital admission was higher using a recall period of 3 months. It can be recommended that lengthened recall periods of 12 months in longitudinal studies are appropriate for gathering self-reported hospital admission in elderly people with myocardial infarction.

6 Acknowledgements

The KORA research platform (KORA, Cooperative Research in the Region of Augsburg) was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. The KORA-Age project was financed by the German Federal Ministry of Education and Research (BMBF FKZ 01ET0713) as part of the ‘Health in old age’ programme. We thank all members of the Helmholtz Zentrum München who were involved in the conduct of the study. Furthermore, we wish to thank the field staff in Augsburg, especially Barbara Dick, Claudia Greschik, Birgit Kling, Dorothea Lukitsch, Sonja Reiserer, Gisela Sietas, and Rupert Wende.

7 References

1. Petrou S, Gray A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. BMJ. 2011;342:d1548. Epub 2011/04/09.
2. Ridyard CH, Hughes DA. Methods for the collection of resource use data within clinical trials: a systematic review of studies funded by the UK Health Technology Assessment program. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2010;13(8):867-72. Epub 2010/10/16.
3. Bhandari A, Wagner T. Self-reported utilization of health care services: improving measurement and accuracy. Med Care Res Rev. 2006;63(2):217-35.
4. Schulz E, Leidl R, Konig HH. The impact of ageing on hospital care and long-term care--the example of Germany. Health Policy. 2004;67(1):57-74. Epub 2004/01/17.
5. Clarke PM, Fiebig DG, Gerdtham UG. Optimal recall length in survey design. Journal of health economics. 2008;27(5):1275-84. Epub 2008/08/01.
6. Kjellsson G, Clarke P, Gerdtham UG. Forgetting to remember or remembering to forget: A study of the recall period length in health care survey questions. Journal of Health Economics. 2014;35:34-46.
7. Seidl H, Meisinger C, Wende R, Holle R. Empirical analysis shows reduced cost data collection may be an efficient method in economic clinical trials. BMC Health Serv Res. 2012;12:318. Epub 2012/09/18.
8. Lubeck DP, Hubert HB. Self-report was a viable method for obtaining health care utilization data in community-dwelling seniors. J Clin Epidemiol. 2005;58(3):286-90.
9. Palmer L, Johnston SS, Rousculp MD, Chu B-C, Nichol KL, Mahadevia PJ. Agreement between Internet-Based Self- and Proxy-Reported Health Care Resource Utilization and Administrative Health Care Claims. Value in Health. 2012;15(3):458-65.
10. Raina P, Torrance-Rynard V, Wong M, Woodward C. Agreement between self-reported and routinely collected health-care utilization data among seniors. Health Serv Res. 2002;37(3):751-74. Epub 2002/07/23.
11. van Dalen MT, Suijker JJ, MacNeil-Vroomen J, van Rijn M, Moll van Charante EP, de Rooij SE, et al. Self-report of healthcare utilization among community-dwelling older persons: a prospective cohort study. PLoS One. 2014;9(4):e93372. Epub 2014/04/09.
12. Wolinsky FD, Miller TR, An H, Geweke JF, Wallace RB, Wright KB, et al. Hospital episodes and physician visits: the concordance between self-reports and medicare claims. Med Care. 2007;45(4):300-7. Epub 2007/05/15.
13. Heinrich S, Deister A, Birker T, Hierholzer C, Weigelt I, Zeichner D, et al. Accuracy of self-reports of mental health care utilization and calculated costs compared to hospital records. Psychiatry Res. 2011;185(1–2):261-8.
14. Rozario PA, Morrow-Howell N, Proctor E. Comparing the congruency of self-report and provider records of depressed elders' service use by provider type. Med Care. 2004;42(10):952-9.
15. Petrou S, Murray L, Cooper P, Davidson LL. The accuracy of self-reported healthcare resource utilization in health economic studies. Int J Technol Assess Health Care. 2002;18(3):705-10.
16. Short ME, Goetzel RZ, Pei X, Tabrizi MJ, Ozminkowski RJ, Gibson TB, et al. How accurate are self-reports? Analysis of self-reported health care utilization and absence when compared with administrative data. J Occup Environ Med. 2009;51(7):786-96. Epub 2009/06/17.
17. Hunger M, Kirchberger I, Holle R, Seidl H, Kuch B, Wende R, et al. Does nurse-based case management for aged myocardial infarction patients improve risk factors, physical functioning and mental health? The KORINNA trial. European journal of preventive cardiology. 2014. Epub 2014/02/14.
18. Kirchberger I, Meisinger C, Seidl H, Wende R, Kuch B, Holle R. Nurse-based case management for aged patients with myocardial infarction: study protocol of a randomized controlled trial. BMC geriatrics. 2010;10:29.
19. Meisinger C, Stollenwerk B, Kirchberger I, Seidl H, Wende R, Kuch B, et al. Effects of a nurse-based case management compared to usual care among aged patients with myocardial infarction: results from the randomized controlled KORINNA study. BMC Geriatr. 2013;13(1):115. Epub 2013/10/31.

20. Seidl H, Hunger M, Leidl R, Meisinger C, Wende R, Kuch B, et al. Cost-effectiveness of nurse-based case management versus usual care for elderly patients with myocardial infarction: results from the KORINNA study. Eur J Health Econ. 2014. Epub 2014/08/12.
21. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12(3):189-98. Epub 1975/11/01.
22. Cacciatore F, Abete P, Ferrara N, Calabrese C, Napoli C, Maggi S, et al. Congestive heart failure and cognitive impairment in an older population. Osservatorio Geriatrico Campano Study Group. J Am Geriatr Soc. 1998;46(11):1343-8. Epub 1998/11/11.
23. Sheikh JI, Yesavage JA. Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. Clin Gerontol. 1986;5:165- 73.
24. Wancata J, Alexandrowicz R, Marquart B, Weiss M, Friedrich F. The criterion validity of the Geriatric Depression Scale: a systematic review. Acta Psychiatr Scand. 2006;114(6):398-410. Epub 2006/11/08.
25. The EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy. 1990;16(3):199-208.
26. Mahoney FI, Barthel DW. Functional Evaluation: The Barthel Index. Md State Med J. 1965;14:61-5.
27. Gibbons RD, Hedeker D, DuToit S. Advances in analysis of longitudinal data. Annual review of clinical psychology. 2010;6:79-107. Epub 2010/03/03.
28. Molenberghs G, Kenward MG. Missing Data in Clinical Studies. UK: John Wiley & Sons Ltd; 2007.
29. Bock JO, Brettschneider C, Seidl H, Bowles D, Holle R, Greiner W, et al. [Calculation of standardised unit costs from a societal perspective for health economic evaluation]. Gesundheitswesen. 2015;77(1):53-61. Epub 2014/07/16. Ermittlung standardisierter Bewertungssatze aus gesellschaftlicher Perspektive fur die gesundheitsokonomische Evaluation.
30. Wallihan DB, Stump TE, Callahan CM. Accuracy of self-reported health services use and patterns of care among urban older adults. Med Care. 1999;37(7):662-70.
31. Yu S-T, Chang H-Y, Lin M-C, Lin Y-H. Agreement between self-reported and health insurance claims on utilization of health care: A population study. J Clin Epidemiol. 2009;62(12):1316-22.