

First assessment of a cost questionnaire for its agreement with administrative databases in females with stress urinary incontinence in Switzerland: a validation study

Céline Moetteli^{a, b*}, Jan Taeymans^{a, c}, Irene Koenig^a

^aBern University of Applied Sciences, Department of Health Professions, Division of Physiotherapy, Bern, Switzerland

^bZurich University of Applied Sciences, School of Health Professions, Institute of Physiotherapy, Winterthur, Switzerland

^cVrije Universiteit Brussel, Faculty of Physical Education and Physiotherapy, Brussels, Belgium

*Corresponding author

E-mail: celine.moetteli@zhaw.ch (CM)

2 Abstract

3 **Introduction:** Economic research in the field of pelvic floor rehabilitation is lacking in
4 transparency of methodology and consideration of indirect costs. Service use questionnaires
5 collect cost data with the broad approach needed for the societal perspective of health
6 economic studies. Nevertheless, these questionnaires are seldom validated. Therefore, this
7 study aimed to evaluate the accuracy of a questionnaire to assess healthcare costs of
8 females with stress urinary incontinence (SUI) in Switzerland.

9 **Methods:** A questionnaire was developed to measure direct medical, direct non-medical and
10 indirect costs in females with stress urinary incontinence living in the larger Bern area in
11 Switzerland. Participants were recruited out of a prior study sample, where they had already
12 completed the cost questionnaire. Administrative databases' cost data were obtained for the
13 investigation period. The agreement between the reported direct medical costs of the
14 questionnaire with those from the providers' billings was assessed using Pearson correlation
15 coefficient (r), intraclass correlation coefficient (ICC) for agreement, Lin's concordance
16 correlation coefficient (LINCCC) and limits of agreement (LOA).

17 **Results:** Fourteen participants were eligible for this study. Agreement between the two
18 methods presented a Pearson's r of 0.64 [95%CI 0.16 to 1.00], ICC of 0.79 [95% CI 0.33 to
19 0.93] and LINCCC 0.75 [95% CI 0.19 to 0.86]. LOA results showed no evidence for
20 systematically under- or overreporting.

21 **Conclusions:** The developed questionnaire seemed to be an acceptable representative for
22 direct medical costs in females with stress urinary incontinence in Switzerland. However,
23 direct non-medical and indirect costs were not compared to an external criterion.

24

25

26 **Keywords:** Urinary Incontinence, Stress; Data Collection; Surveys and Questionnaires;
27 Health Care Costs

28 Introduction

29 In Switzerland, per capita healthcare expenditures rise steadily from year to year. In 2020,
30 healthcare costs amounted 11.8% of the gross domestic product (1). In times of increasing
31 health expenditures, it is indispensable to perform economic evaluations to ensure cost-
32 effective resource allocation in healthcare organizations.

33 The estimation of costs is an essential aspect when conducting health economic analyses.
34 Thus, valid tools are required to obtain useful cost data. Self-reported questionnaires are one
35 alternative to gain resource utilization data (2). The societal perspective is recommended for
36 health economic analyses, since costs shifting between sectors might be detected (3).
37 Hence, not only direct medical costs are taken into account, but all arising expenses
38 concerning the macroeconomic situation (4). In contrast to health insurance administrative
39 data, self-reported questionnaires allow this broad capture of health related costs (5). Apart
40 from direct medical costs, direct non-medical costs such as travel expenses, costs resulting
41 from paid or unpaid informal care or indirect costs such as absenteeism, presenteeism or
42 early retirement are considered (6). Economic estimations based on self-reported
43 questionnaires are, unlike medical records, not limited by lacking cooperation of
44 administrative staff or restrictions due to privacy and security concerns. Although
45 administrative databases only allow insight into direct medical costs, they are a rich source of
46 accurate data. While retrospective patient reported questionnaires may be affected by recall
47 bias, administrative data does not suffer from such influences (5). However, administrative
48 data may be inaccessible in a time- and cost-efficient manner and inappropriate due to their
49 limited perspective (2).

50 Therefore, within the scope of a cost of illness (COI) study, a questionnaire to measure
51 health related costs in females with stress urinary incontinence (SUI) was developed. Urinary
52 incontinence is a common problem among females, causing almost ten billion Euros of direct
53 and indirect costs in Europe per year (7). Stress urinary incontinence is the most common
54 type of this condition with a mean prevalence of 35% in Europe (8). In the field of pelvic floor

55 medicine, transparency in the costing methodology and consideration of a broad perspective
56 of costs are lacking globally as well as in Switzerland (7). Thus, the aforementioned
57 questionnaire should be eligible for further health economic evaluations in urinary
58 incontinence. Therefore, it ought to be published on “Database of Instruments for Resource
59 Use Measurement (DIRUM)”; an open-access database of resource-use questionnaires for
60 use by health economists (9).

61 **Methods**

62 The aim of this study was to evaluate the questionnaire’s accuracy of assessing health
63 related costs in females with stress urinary incontinence living in the larger area around Bern
64 in Switzerland. Since there is no standardized method for validating health economic
65 questionnaires, health insurance administrative data was used as the comparative value as
66 seen in previous papers (5,6,10,11).

67 **Questionnaire development**

68 The questionnaire was developed for a previous COI analysis in patients with stress urinary
69 incontinence. However, validity of the data gathered by this tool was never evaluated. The
70 “Health status, comorbidities and cost-of-illness in females with stress urinary incontinence
71 living in the Canton of Bern” (12) was embedded in a randomized controlled trial investigating
72 the effects of two pelvic floor muscle training protocols (13). The first version of the
73 instrument was developed based on a questionnaire that has already been used in a study
74 from the Universities of Ghent and Antwerp (Belgium), which evaluated health status and
75 costs of obesity in adult females in Flanders (14). Face validity of the adapted cost
76 questionnaire was operationalized by comprehensibility and completeness of its items and
77 assessed by six experts in the field of pelvic floor rehabilitation. Additionally, the
78 questionnaire was confirmed for comprehensibility by two pilot-test subjects. To review
79 criterion validity in a first point, these two test subjects filled out the cost questionnaire
80 retrospectively for the last four months and provided the health insurance bills for the same
81 period of observation. The comparison with the survey data did not raise any major

82 concerns, or the need for further modifications (12).

83 The present cost questionnaire under investigation assessed participants' characteristics,
84 health status and health related costs retrospectively over four months. This included all
85 direct medical costs such as medical and therapeutic treatment costs, raises for dental
86 services, complementary medicine, and medications. Consumed service units and price units
87 were asked separately. In addition, direct non-medical costs (e.g., incontinence aids) and
88 indirect costs (e.g., productivity loss such as work absenteeism and presenteeism) were
89 assessed. "Filter questions" were used as guiding assistance for costs reporting (6). The
90 questionnaire in full version can be found under "additional files".

91 **Setting**

92 The previous COI study of Koenig et al. (12) used a prevalence based, bottom-up approach.
93 According to a prevalence based approach guidelines, questions of the survey targeted not
94 only expenditures due to SUI but the entirety of healthcare expenditure (15). Out of 96
95 participants in the original randomized controlled trial (RCT), 37 adult female participants
96 volunteered in the COI study and completed the cost questionnaire in an individual four-
97 month time horizon between 2016 and 2017. Each participant returned her questionnaire
98 during a post-intervention measurement for the RCT. The answers were then checked by the
99 responsible examiner for completeness and missing data were completed immediately in the
100 presence of the participants. Based on unit prices, Koenig et al. (12) estimated the total
101 healthcare costs of the investigated four months.

102 The Swiss healthcare system is highly decentralized. Compulsory health insurance is offered
103 by competing insurers with minimal annual deductibles of 300 Swiss Francs (CHF) for
104 insured adults. Insured persons may increase deductibles up to 2'500 CHF. Insured persons
105 also pay 10% coinsurance of these services with an annual limit of 700 CHF for adults (16).

106 Residents are free to take out a supplementary insurance. In contrast to the compulsory
107 health insurance, coverage of services may differ between insurance providers.

108 Supplementary insurances might include, for example, complementary medicine, dental

109 treatment, or health prevention services. There are different methods of charging the fee:
110 service providers can either invoice the insurances directly or they can bill patients, which
111 settle up the account and claim reimbursement from the insurer (16).

112 Study participants and procedure

113 Each participant received written information about the study design and procedures and
114 signed the informed consent form between October 2019 and February 2020. With
115 volunteer's power of attorney, one of the authors (CM) was allowed insight into personal
116 health insurances billing data of the investigated four-month period between 2016 and 2017.

117 The inclusion criteria for the previous randomized controlled trial were: age range 18 to 60
118 years, body mass index (BMI) between 18 to 30 kg/m², nulliparous or at least 12 months
119 postpartum with a negative pregnancy test, ability to read and understand the German
120 language and to run on a treadmill. Exclusion criteria were: acute urinary tract or vaginal
121 infections, urogenital surgery, predominant overactive bladder and urogenital prolapse over
122 grade one pelvic organ prolapse quantification system (POP-Q). These criteria met the
123 International Consultation on Incontinence's recommendation for diagnosis of SUI (17).

124 Statistical analysis

125 Characteristics of the sample were compared to the original cohort of the RCT with a
126 Student's *t*-test for independent samples. As a reference for the degree of incontinence, the
127 Oxford Grading Scale was used. This scale is a classification system of the pelvic floor
128 muscle strength where zero represents no palpable contraction and the maximum score of
129 five signifies strong contraction (18). Median and interquartile range (IQR) of observed
130 muscle strength values were calculated for the original and the current samples.

131 For comparability with other studies (19), descriptive parametric statistics were used to
132 analyze the cost data, despite the probability of a skewed distribution. The items "medical
133 care" and "physiotherapy" fell into the scope of compulsory health insurance. The items
134 "dental care", "complementary treatment" and "medication" were categorized as services
135 within supplementary insurance. To measure criterion validity, the differences between

136 reported costs and health insurance billings were calculated and compared with “0”
137 difference using a one-sample *t*-test. For a first visual impression of the relationship between
138 the datasets, a scatterplot was performed. Cost data is usually skewed due to, for example,
139 medical complications. Therefore, data was log transformed for approximate normality to
140 conduct calculations of correlation (20,21).

141 As a preliminary test, consistency of variables was indicated by Pearson’s correlation
142 coefficient (*r*). Agreement was measured using intraclass correlation coefficient (ICC). Since
143 the variables were continuous and reported in the same units, ICC was appropriate (22). ICC
144 two-way, mixed model formula for agreement was applied. As a reference value for small
145 sample sizes, Lin’s concordance correlation coefficient (LINCCC) was added (23).
146 Pearson’s *r*, ICC and LINCCC are measurements of correlation ranging from 1 (perfect
147 agreement) to -1 (perfect inverse agreement) (23,24). The following levels for interpreting
148 accuracy were applied: “poor” = < 0.40, “fair to good” = 0.40 to 0.74, “excellent” = 0.75 to
149 1.00 (25). Standard error of measurement (SEM) was calculated (26). Systematic bias was
150 assessed using limits of agreement (LOA) and were presented in a Bland Altman plot (6).
151 The visual result was supported by the calculation of linear regression of the difference
152 between the costs over mean value of the costs. The procedure was doublechecked by two
153 of the authors independently.

154 All data analyses were performed using SPSS (IBM Corp. Released 2017. IBM SPSS
155 Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.) and with a macro written in
156 Excel for the calculation of LINCCC. Significance was set at 5% level of error.

157 Results

158 Out of the 37 participants of the COI study, 14 persons volunteered in the present agreement
159 study. The participants were insured in eight different healthcare insurance companies,
160 which all provided detailed billing data after written request from the participants.

161

162 Table 1 shows the characteristics of the participants such as age, BMI, number of vaginal
163 births as well as sectio and scores on the Oxford Grading Scale. The present sample differed
164 in none of these variables significantly from the original RCT population ($p > 0.05$).

Table 1 Comparison of previous and current sample's participants characteristics

	Present study (n=14)		RCT (n=96)		<i>p</i>
	Mean	SD	Mean	SD	
Age (years)	54.0	9.4	51.0	9.9	0.290
BMI (kg/m ²)	24.8	4.6	24.8	3.8	0.993
No vaginal birth	1.20	0.90	1.60	1.10	0.208
No sectio birth	0.15	0.38	0.30	0.70	0.458
Oxford Grading Scale (0-5)	Median: 3	IQR: 1	Median: 3	IQR: 1	-

SD: Standard deviation; *IQR*: Interquartile range

165
166 In table 2, data was compiled by subgroups. Mean values and standard deviations (SD) for
167 reported costs and billings were opposed for the items “medical care” and “physiotherapy”
168 indicating compulsory services, and “dental care”, “complementary treatment” and
169 “medication” representing supplementary services.

Table 2 Subgroup means and SDs of reported costs and billings

	Reported cost (CHF)		Billings (CHF)	
	Mean	SD	Mean	SD
Compulsory services				
Medical care	513.60	594.90	523.80	854.90
Physiotherapy	508.50	130.70	481.35	116.45
Supplementary services				
Dentist	70.50	116.90	23.95	64.85
Complementary treatment	126.10	248.20	195.00	397.40
Medication	69.85	108.15	59.45	143.75

CHF: Swiss Francs; *SD*: Standard deviation

171
172 Table 3 depicts the analysis of correlation. Calculations were conducted for overall costs as
173 well as expenditures of compulsory and supplementary services separately. Total sum of
174 reported costs during the four-months period was CHF 18'038.85 and CHF 17'113. 85 for
175 billings. Overall correlation between the reported costs and the billing sums yielded in a

176 Pearson's r of 0.64 [95% CI 0.16 to 0], ICC of 0.79 [95% CI 0.33 to 0.93] and LINCCC of
177 0.75 [95% CI 0.19 to 0.86].

178

179

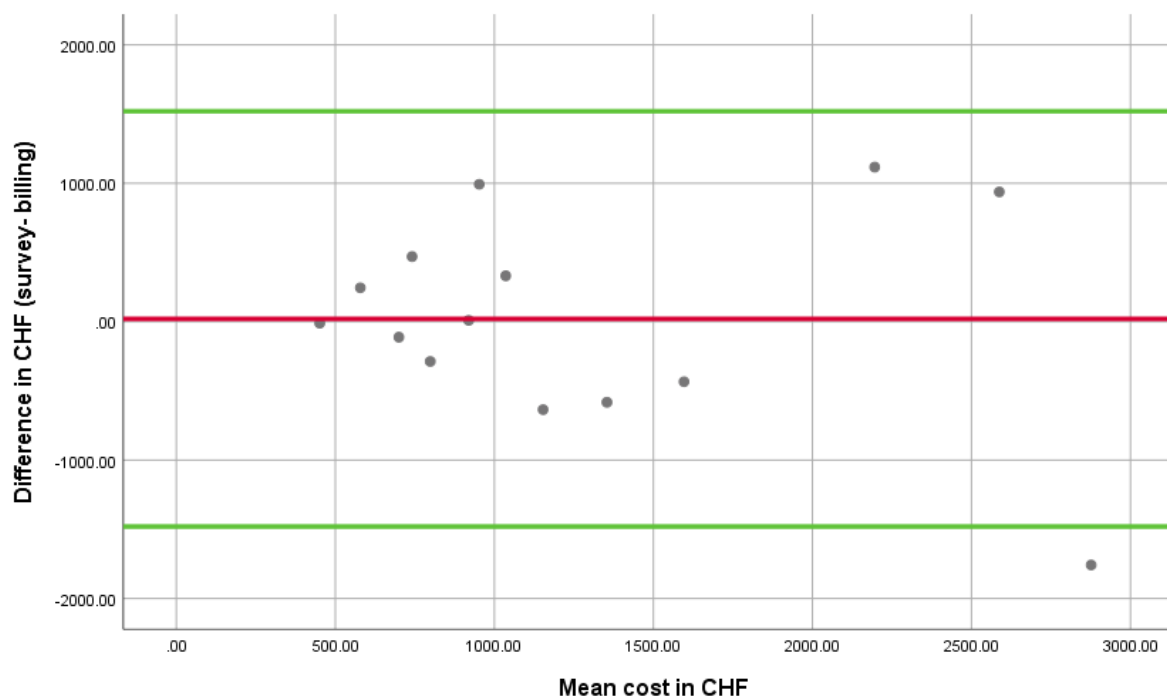
Table 3 Sum, difference and correlations between reported costs and billings

	Sum CHF	Diff. CHF	<i>p</i>	Pearson's <i>r</i> [95% CI]	ICC [95% CI]	LINCCC [95% CI]	SEM CHF
Total reported cost	18'038.85			0.64	0.79	0.75	
Total billings	17'773.85	265.00	0.928	[0.16-1]	[0.33-0.93]	[0.19-0.86]	354.30
Compulsory services reported	14'309.35			0.61	0.77	0.71	
Compulsory services billing	14'071.55	237.80	0.932	[0.12-1]	[0.27-0.93]	[0.14-0.85]	352.95
Supplementary service reported	3'729.50			0.48	0.65	0.49	
Supplementary service billing	3'702.30	27.20	0.979	[-0.07-1]	[-0.07-0.89]	[0.05-0.78]	156.95

CHF: Swiss Francs; *Diff*: Difference; *CI*: Confidence interval; *ICC*: Intraclass correlation coefficient; *LINCCC*: Lin's concordance correlation coefficient; *SEM*: Standard error of measurement

180 Systematic errors of overall data are presented in the Bland Altman Plot in Figure 1. Linear
181 regression from cost difference over mean costs yielded a beta regression coefficient of
182 -0.17 ($p = 0.564$).

183 **Figure 1 Bland Altman Plot**



184
185 **Figure 1 Bland Altman Plot.** Mean values of provider and survey data displayed in the x-
186 axis are plotted against the difference between the values in the y-axis. Limits of agreement
187 are represented in green. The red line is the mean of difference and is drawn at 20.3

188 Discussion

189 Assessment of validity

190 This study aimed to validate a questionnaire assessing costs in females with stress urinary
191 incontinence living in the larger area around Bern in Switzerland. Although analysis of
192 correlation is known as a reliability process, it is also used as a method to determine criterion
193 validity between the measurement instrument and the gold standard (22). As a gold
194 standard, health insurance billing was used in this study.

195 Students' *t*-test for independent samples showed no significant difference in demographic
196 data of the 14 participants enrolled in the current study to the original cohort of 96 females.
197 Since the participants of the RCT were diagnosed according to the criteria of the

198 International Consultation on Incontinence, the current sample was representative for
199 females with stress urinary incontinence.

200 Healthcare expenditures analyzed by the items “medical care”, “physiotherapy”, “dental
201 care”, “complementary treatment” and “medications” showed large standard deviations in
202 both reported and administrative data. Since each of those items represented different
203 medical services, this variability seemed to have occurred naturally. For example, a
204 respondent could accrue a ticket for a simple laboratory service of about CHF 20 in the
205 category of medical care as well as a surgical procedure of CHF 1’500. Despite the observed
206 variability, when comparing mean values of reported costs and billing data, subgroups
207 appeared similar. This was particularly relevant for the items “medical care” and
208 “physiotherapy”. Participants seemed to estimate costs for these services accurately and
209 consistently with insurance billings. The differences of datasets get larger, when it comes to
210 the costs of dental care and complementary treatments. Reported dentist costs were three
211 times higher than insurances recorded billings. Since Swiss residents are free to take out a
212 supplementary insurance of their needs, different services might be covered or not. If
213 participants had a supplementary dental health insurance, dentists’ costs would be covered
214 by the insurance. If not, residents would most likely settle the bill out of pocket without
215 submitting the invoice to the insurance. Consequently, companies might be incompletely
216 informed and reflect the true costs insufficiently. Thus, the patient reported questionnaire
217 could be more accurate than the so-called gold standard itself. Concerning complementary
218 treatments, billings were higher than reported costs. Since overestimation of billing data is
219 unlikely (5), underestimation of patient reported cost must be assumed. Pinto et al. (5)
220 observed this behavior of underreporting regarding doctor visits. Recall bias side lined,
221 difference of reported costs may also be a result of the “reverse telescoping” effect.
222 Telescoping occurs, when participants unconsciously lengthen their period of investigation
223 and report events erroneously into the recall time. In contrast, treatments within the time
224 frame may be “reverse telescoped” out of the recall period (27). True costs might even be
225 higher than insurance’ billings revealed. Depending on the coverage of services, participants

226 may or may not have submitted the invoices to the companies. Therefore, unsubmitted
227 billings must be assumed, which would lead to a simultaneous underestimation of insurances
228 expenditures for complementary treatments. In the case of medications, insurance
229 companies' information is generally restricted to prescribed drugs. Whereas the patient
230 reported cost questionnaire captured all types of medication, including out of pocket
231 utilizations. Overall sums of reported costs and billings were similar but with large standard
232 deviations. Subsequently, the data range had a similar impact on each arm of the distribution
233 and averaged itself out. Correspondingly similar values of reported costs were found after
234 dividing the costs into categories of compulsory and supplementary services. Differences
235 between the data sets were nonsignificant in all the three calculations. Therefore, Pearson's r
236 correlation coefficient was calculated. In contrast to ICC agreement, Pearson's r does not
237 take systematic error into account (22). Whereas ICC measures the degree to which one
238 variable can be equated to another variable, Pearson's r assesses the linearity of the
239 relationship between the two variables. Lin's concordance correlation coefficient is known for
240 its robustness even with small sample sizes (23,24). It evaluates the degree to which pairs
241 fall on the 45° line when plotted (28). In this present study, variability between the degree of
242 correlation between the three methods occurred. Given the small number of samples,
243 Pearson's r and ICC may have yielded unstable calculations. Hence, ICC reflected a higher
244 correlation than Pearson's r despite the more restrictive formula (22). Point estimations of
245 correlation for overall data were "fair/ good" to "excellent" according to the levels of accuracy
246 with a Pearson's r of 0.54 [95% CI 0.16 to 1.00], ICC of 0.79 [95% CI 0.33 to 0.93] and
247 LINCCC of 0.75 [95% CI 0.19 to 0.86]. Nevertheless, the large confidence intervals indicated
248 an ambiguous result. As expected, due to the descriptive values, correlations of
249 supplementary services were lower than compulsory services with point estimations rated
250 "fair to good" for supplementary services respectively, "fair to good" and "excellent" for
251 compulsory services.

252 The Bland Altman plot (figure 1) depicts the mean value of the pairs against differences
253 between the pairs. Each dot on the plot represents data from one participant. There are six

254 dots each above and below the mean visible. The limits of agreement are wide and thus
255 consistent with the large confidence intervals of the correlations. Out of 14 dots, 13 are
256 located within the limits of agreement and one is situated exteriorly. There is a trend
257 noticeable, where the more the mean increases, the larger the differences become. This
258 finding suggests that participants who have less health service consumption were able to
259 estimate their costs better than those with much healthcare consumption. The calculation of
260 the linear regression confirmed the first impression of no evident systematic bias. Beta
261 coefficient was -0.17 ($p = 0.564$), indicating that there was no proportional bias involved.
262 Indeed, the plot shows that the deviations of the mean were evenly balanced between under-
263 and overreporting of study participants. In the systematic review of Leggett et al. (2), a
264 tendency of underreporting was noticed. The 15 included self-reported questionnaires
265 seemed to capture the number of hospitalizations and emergency room visits accurately,
266 whereas participants tended to underestimate the number of visits to a general practitioner
267 (GP).

268 Pinto et al. (5) developed a questionnaire to capture healthcare use and costs in patients
269 with osteoarthritis in New Zealand. This study included 50 participants. The reported costs of
270 the three-month investigation period were compared to administrative data. GP services,
271 hospital services and medication use were collated. According to the authors, correlations
272 were acceptable concerning GP services and hospitalizations (LINCCC >0.4) but exhibited
273 poor agreement for medication (LINCCC < 0.4). These results were in line with findings of
274 the current included females with SUI. Depending on the cost categories, concordance
275 between survey and administrative data corresponded well and the questionnaire seemed to
276 give an appropriate reflection of direct medical costs. Nonetheless, administrative databases
277 were limited as gold standard, since they captured a more restricted perspective of costs
278 than the patient reported questionnaires.

279 This study attempted an important step for standardized economic research in the field of
280 pelvic floor medicine and rehabilitation. Further research is needed to evaluate and

281 strengthen the results of this present study on a greater sample size. Since the cost
282 questionnaire is only available on paper, a digital version is required. For this future version,
283 a feasibility study will be needed.

284 Limitations

285 In their systematic review, Leggett et al. (2) showed that few studies have aimed to measure
286 the accuracy of healthcare service use and cost data questionnaires. Thus, no unambiguous
287 method is known (29) and limitations need to be declared. The first obstacle for assessing
288 accuracy of such questionnaire is the lack of an appropriate “gold standard” (29). Insurance
289 companies do not cover all types of healthcare costs and are therefore not informed about
290 the totality of expenses. Due to the decentralized Swiss healthcare system, there is no single
291 omniscient administrative database. Survey data were thus compared to a deficient criterion.
292 Secondly, since there is no appropriate comparator for a broad perspective of costs, the cost
293 questionnaire was assessed for a subset of questions only. Hence, declaration of validity
294 was only possible for the part of direct medical costs. Direct non-medical and indirect costs
295 were neglected. Thirdly, sample size of the present study was low ($n = 14$). Due to the fact,
296 that the study population was originally participating in an RCT and was committed to the
297 following COI analysis, only few people agreed with another observation. Consequently, the
298 low number of participants may have generated unstable calculations. An unambiguous
299 result would have been achieved by a greater cohort.

300 Conclusion

301 The developed questionnaire for measuring costs in females with stress urinary incontinence
302 in Switzerland seemed to capture direct medical costs sufficiently. Overall correlation of
303 patient reported data and health insurance billings yielded in a Pearson's r of 0.64 [95% CI
304 0.16 to 1], ICC of 0.79 [95% CI 0.33 to 0.93] and LINCCC of 0.75 [95% CI 0.19 to 0.86].
305 Compulsory services such as medical care and physiotherapy corresponded better with
306 administrative database than supplementary services such as expenses for dentist,

307 complementary treatment, and medication. No proportional bias of under- or overreporting

308 occurred.

309

310 List of abbreviations

BMI	body mass index
CHF	Swiss Franks
CI	confidence interval
COI	cost of illness
DIRUM	Database of Instruments for Resource Use Measurement
GP	general practitioner
ICC	intraclass correlation coefficient
IQR	interquartile range
LINCCC	Lin's concordance correlation coefficient
LOA	limits of agreement
POP-Q	pelvic organ prolapse quantification system
RCT	randomized controlled trial
SD	standard deviations
SEM	standard error of measurement
SUI	stress urinary incontinence

311

312 Declarations

313 Ethics approval and consent to participate

314 The study protocol has been approved by the Ethics Committee of the Canton of Bern,
315 Switzerland (reference number 249/14 on 27 June 2019) in accordance with the Declaration
316 of Helsinki and the Swiss Human Research Act. Participants received written information
317 about data collection and data protection as well as signing an informed consent form.

318 Authors' contributions

319 CM, IK, and JT provided input into the design of the study. CM obtained informed consent
320 from study participants and gathered cost data from health insurers. CM and JT analyzed
321 and interpreted the data. CM prepared the first draft of the manuscript. IK and JT reviewed
322 the manuscript and provided input. All authors read and approved the final version of the
323 manuscript.

324 Acknowledgements

325 The authors would like to thank Sara Mahnig for her methodological advice, André Meichtry
326 for his statistical support and Ben Tite for proofreading this article. Thanks also to the
327 employees of responsible insurance companies, which compiled requested data.

328 **Trial registration:** Baseline randomized controlled trial was registered at clinicaltrials.gov
329 (www.clinicaltrials.gov/ study identifier: NCT02318251) in February 2018.

330

331 References

- 332 1. OECD. (24. Oktober, 2022). Anteil der Gesundheitsausgaben in der Schweiz am
333 Bruttoinlandsprodukt von 2010 bis 2020 [Graph]. In Statista. Zugriff am 21. März 2023,
334 von [https://de.statista.com/statistik/daten/studie/18342/umfrage/anteil-der-](https://de.statista.com/statistik/daten/studie/18342/umfrage/anteil-der-gesundheitsausgaben-in-der-schweiz-am-bruttoinlandsprodukt-bip/)
335 [gesundheitsausgaben-in-der-schweiz-am-bruttoinlandsprodukt-bip/](https://de.statista.com/statistik/daten/studie/18342/umfrage/anteil-der-gesundheitsausgaben-in-der-schweiz-am-bruttoinlandsprodukt-bip/)
- 336 2. Leggett LE, Khadaroo RG, Holroyd-Leduc J, Lorenzetti DL, Hanson H, Wagg A, et al.
337 Measuring Resource Utilization: A **Systematic** Review of Validated Self-Reported
338 Questionnaires. *Medicine*. 2016 Mar;95(10):e2759.
- 339 3. Byford S, Raftery J. Economics notes: Perspectives in economic evaluation. *BMJ*. 1998
340 May 16;316(7143):1529–30.
- 341 4. Schöffski O, Graf von der Schulenburg JM, editors. *Gesundheitsökonomische*
342 *Evaluationen*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012.
- 343 5. Pinto D, Robertson MC, Hansen P, Abbott JH. Good agreement between questionnaire
344 and administrative databases for health care use and costs in patients with osteoarthritis.
345 *BMC Med Res Methodol*. 2011 Dec;11(1):45.
- 346 6. Schweikert B, Hahmann H, Leidl R. Development and first assessment of a
347 questionnaire for health care utilization and costs for cardiac patients. *BMC Health Serv*
348 *Res*. 2008 Dec;8(1):187.
- 349 7. Zwolsman S, Kastelein A, Daams J, Roovers JP, Opmeer BC. Heterogeneity of cost
350 estimates in health economic evaluation research. A systematic review of stress urinary
351 incontinence studies. *Int Urogynecol J*. 2019 Jul;30(7):1045–59.
- 352 8. Hunskaar S, Lose G, Sykes D, Voss S. The prevalence of urinary incontinence in women
353 in four European countries. *BJU Int*. 2004 Feb;93(3):324–30.
- 354 9. DIRUM - Database of Instruments for Resource Use Measurement [Internet]. [cited 2020
355 May 8]. Available from: <http://www.dirum.org/>
- 356 10. Lairson DR, Basu R, Begley CE, Reynolds T. Concordance of survey and billing data in a
357 study of outpatient healthcare cost and utilization among epilepsy patients. *Epilepsy*
358 *Research*. 2009 Nov;87(1):59–69.
- 359 11. Goossens MEJB, Mólken MPMHR van, Vlaeyen JWS, van der Linden SMJP. The cost
360 diary. *Journal of Clinical Epidemiology*. 2000 Jul;53(7):688–95.
- 361 12. Koenig I, Moetteli C, Luginbuehl H, Kuhn A, Taeymans J. Health status, comorbidities
362 and cost-of-illness in females with stress urinary incontinence living in the Canton of
363 Bern. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*. 2020
364 Apr;(150–152):73–9.
- 365 13. Luginbuehl H, Lehmann C, Baeyens JP, Kuhn A, Radlinger L. Involuntary reflexive pelvic
366 floor muscle training in addition to standard training versus standard training alone for
367 women with stress urinary incontinence: study protocol for a randomized controlled trial.
368 *Trials*. 2015 Dec;16(1):524.
- 369 14. Hens W, Vissers D, Annemans L, Gielen J, Van Gaal L, Taeymans J, et al. Health-
370 related costs in a sample of premenopausal non-diabetic overweight or obese females in
371 Antwerp region: a cost-of-illness analysis. *Arch Public Health*. 2018 Dec;76(1):42.

- 372 15. Mauskopf J. Prevalence-based economic evaluation. *Value In Health*. 1998 Nov;Volume
373 1(Issue 4):Pages 251-259.
- 374 16. Mossialos E, Wenzl M, Osborn R, Sarnak D. *International Profiles of Health Care*
375 *Systems*, 2015. The Commonwealth Fund; 2016 Jan.
- 376 17. Avery K, Donovan J, Peters TJ, Shaw C, Gotoh M, Abrams P. ICIQ: A brief and robust
377 measure for evaluating the symptoms and impact of urinary incontinence. *Neurourol*
378 *Urodyn*. 2004;23(4):322–30.
- 379 18. Bo K, Finckenhagen HB. Vaginal palpation of pelvic floor muscle strength: inter-test
380 reproducibility and comparison between palpation and vaginal squeeze pressure. *Acta*
381 *Obstet Gynecol Scand*. 2001 Oct;80(10):883–7.
- 382 19. Chernyak N, Jülich F, Kasperidus J, Stephan A, Begun A, Kalthener M, et al. Time cost
383 of diabetes: Development of a questionnaire to assess time spent on diabetes self-care.
384 *Journal of Diabetes and its Complications*. 2017 Jan;31(1):260–6.
- 385 20. Thompson SG. How should cost data in pragmatic randomised trials be analysed? *BMJ*.
386 2000 Apr 29;320(7243):1197–200.
- 387 21. Euser AM, Dekker FW, le Cessie S. A practical approach to Bland-Altman plots and
388 variation coefficients for log transformed variables. *Journal of Clinical Epidemiology*.
389 2008 Oct;61(10):978–82.
- 390 22. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical*
391 *Guide*. Cambridge: Cambridge University Press; 2011.
- 392 23. Lin LIK. Assay Validation Using the Concordance Correlation Coefficient. *Biometrics*.
393 1992 Jun;48(2):599.
- 394 24. McAlinden C, Khadka J, Pesudovs K. Statistical methods for conducting agreement
395 (comparison of clinical tests) and precision (repeatability or reproducibility) studies in
396 optometry and ophthalmology: Agreement studies. *Ophthalmic and Physiological Optics*.
397 2011 Jul;31(4):330–8.
- 398 25. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions* [Internet].
399 Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2003 [cited 2020 May 8]. (Shewart WA,
400 Wilks SS, editors. *Wiley Series in Probability and Statistics*). Available from:
401 <http://doi.wiley.com/10.1002/0471445428>
- 402 26. Harvill LM. An NCME Instructional Module on. Standard Error of Measurement.
403 *Educational Measure: Issues Practice*. 1991 Jun;10(2):33–41.
- 404 27. Bhandari A, Wagner T. Self-Reported Utilization of Health Care Services: Improving
405 Measurement and Accuracy. *Med Care Res Rev*. 2006 Apr;63(2):217–35.
- 406 28. Lin LIK. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*.
407 1989 Mar;45(1):255.
- 408 29. Byford S, Leese M, Knapp M, Seivewright H, Cameron S, Jones V, et al. Comparison of
409 alternative methods of collection of service use data for the economic evaluation of
410 health care interventions. *Health Econ*. 2007 May;16(5):531–6.

412 Figure Legends

413 **Figure 1.** Bland Altman Plot

414 Supporting information

415 **S1 File** cost questionnaire “Health and cost evaluation in SUI”.