Relationship Between Time Consumption and Quality of Responses to Drug-related Queries: A Study From Seven Drug Information Centers in Scandinavia

Linda Amundstuen Reppe, MSc.Pharm[1,2,3]; Stian Lydersen, PhD[4];
Jan Schjøtt, MD, PhD[5,6,7]; Per Damkier, MD, PhD[8];
Hanne Rolighed Christensen, MD, PhD[9]; Jens Peter Kampmann, MD, Dr.Med.Sci[9];
Ylva Böttiger, MD, PhD[10]; and Olav Spigset, MD, PhD[2,3]

[1]Faculty of Health Sciences, Nord University, Steinkjer, Norway; [2]Department of Laboratory Medicine, Children's and Women's Health, Norwegian University of Science and Technology, Trondheim, Norway; [3]Regional Medicines Information and Pharmacovigilance Center (Midt-Norge), St. Olavs Hospital, Trondheim, Norway; [4]Regional Center for Child and Youth Mental Health and Child Welfare—Central Norway, Norwegian University of Science and Technology, Trondheim, Norway; [5]Section of Clinical Pharmacology, Laboratory of Clinical Biochemistry, Haukeland University Hospital, Bergen, Norway; [6]Department of Clinical Science, Faculty of Medicine and Dentistry, University of Bergen, Bergen, Norway; [7]Regional Medicines Information and Pharmacovigilance Center (Vest), Haukeland University Hospital, Bergen, Norway; [8]Department of Clinical Chemistry & Pharmacology, Odense University Hospital, Odense, Denmark; [9]Department of Clinical Pharmacology, Bispebjerg and Frederiksberg University Hospital, Copenhagen, Denmark; and [10]Clinical Pharmacology, Department of Drug Research, Linköping University, Linköping, Sweden

ABSTRACT

Purpose: The aims of this study were to assess the quality of responses produced by drug information centers (DICs) in Scandinavia, and to study the association between time consumption processing queries and the quality of the responses.

Methods: We posed six identical drug-related queries to seven DICs in Scandinavia, and the time consumption required for processing them was estimated. Clinical pharmacologists (*internal experts*) and general practitioners (*external experts*) reviewed responses individually. We used mixed model linear regression analyses to study the associations between time consumption on one hand and the summarized quality scores and the overall impression of the responses on the other hand.

Findings: Both expert groups generally assessed the quality of the responses as "satisfactory" to "good." A few responses were criticized for being poorly synthesized and less relevant, of which none were quality-assured using co-signatures. For external experts, an increase in time consumption was statistically significantly associated with a decrease in common quality score (change in score, −0.20 per hour of work; 95% CI, −0.33 to −0.06; P = 0.004), and overall impression (change in score, −0.05 per hour of work; 95% CI, −0.08 to −0.01; P = 0.005). No such associations were found for the internal experts' assessment.

Implications: To our knowledge, this is the first study of the association between time consumption and quality of responses to drug-related queries in DICs. The quality of responses were in general good, but time consumption and quality were only weakly associated in this setting. (*Clin Ther.* 2016;38:1738–1749) © 2016 The Authors. Published by Elsevier HS Journals, Inc.

Key words: drug information services, physicians, "quality assurance, health care", time factors.

INTRODUCTION

There are no established accepted methods or criteria for measuring the quality of drug information centers' (DICs') responses to queries.[1] From the DICs' points of view, relevant quality assurance may include properly trained staff members, standardized working procedures, documentation of the working process,[2] and the use of co-signature by another staff member.[3] Most studies assessing the quality of drug information services have been designed as user satisfaction surveys, aiming to address health care professionals' evaluations of the responses.[4–7] Such surveys have been criticized as lacking objectivity[8,9] and as being biased by evaluation of one's own center and not including users not responding.[9] Some user surveys have been focusing on the impact of these services on patient care, assessed by health care professionals.[4–6,10–12] These methods have retrospective approaches, and the lack of controls to which the actual outcomes could be compared has been criticized.[9] In addition, many factors other than the DIC's response can affect patient outcome.[13] Scandinavian DICs have also published results from users' surveys, and the users have generally been very satisfied with the services.[4–6,14]

Regardless of outcome measures, such surveys do not indicate where the strengths and weaknesses of DICs' responses lie.[2] In order to ensure high quality of written responses from DICs, another proposed method of assessment is to use an external committee to review them.[9,13] Several studies have included such external reviews.[8,15,16] Yet another way of measuring the quality of responses from DICs has been to pose the same query to several DICs at the same time, comparing the responses given from different services to each other and/or to a "control response" giving the correct answer.[15–19]

Previous studies aiming to compare responses from different DICs to identical queries have generally revealed unsatisfactory results.[15,17–19] Halbert et al[19] posed the same telephone query to 90 different US DICs in 1977. Ten centers were not able to identify the drug in question, and 22 centers provided information that was judged to be less than adequate. Gallo et al[15] posed identical queries to 20 hospital-based US DICs. A panel of five clinical pharmacists assessed the directness, applicability, accuracy, and completeness of the answers. Only nine DICs provided an answer. The highest possible score

was 100, and the responses' scores ranged from 23 to 84, with a median of 62.

In 1990, Beaird et al[18] randomly selected 59 of 154 DICs in the United States. They performed a telephone request requiring the identification of didanosine. If the center was able to identify the drug, the staff member was presented with a patient case with symptoms of acute pancreatitis. Of the 56 centers that were successfully contacted, only 16 identified the drug as didanosine, and 4 recognized the clinical symptoms of pancreatitis and associated it with the use of didanosine. Calis et al[17] evaluated responses from US DICs responding to four drug-related queries. Of the 79 centers that responded to all four queries, none provided a correct overall response to all, 13 had three overall correct responses, 42 had two overall correct responses, 21 had one correct overall response; 3 centers failed to answer any of the queries correctly.

Better results were reported from two literature search services in Australia serving general practitioners (GPs). The services focused on answering queries requiring thorough searching for evidence-based documentation. Both services answered the same 14 queries asked during the study period. One person with experience in evidence-based medicine rated the concordance between the reports. There were substantial intersite differences in the evidence sections of four of the reports, and minor differences in another four. There were, however, no substantial differences in the overall conclusions of the reports.[16]

The Scandinavian DICs provide written responses to almost all drug-related queries posed to the centers. The centers are quite similar in structure and types of queries, and have recently been studied in terms of time consumption when responding to drug queries.[20] That study revealed that time spent by staff members processing queries (in the present article designated *time consumption*) varied largely both between queries and between DICs. The quality of written responses from these centers has not previously been compared.

One aim of this study was to assess the quality of responses processed by Scandinavian DICs using both internal experts (clinical pharmacologists) and external experts (GPs). Another aim was to investigate whether there was an association between time consumed when processing the responses and their quality.

MATERIALS AND METHODS

All regional Scandinavian DICs (n = 11) were invited to participate in the study. These DICs are governmentally funded and affiliated with clinical pharmacology units at university hospitals in their respective geographic regions. The centers are manned by pharmacists and clinical pharmacologists, and the centers' aim is to contribute to the rational use of drugs. One of the main activities is to respond to health care professionals' drug-related queries. Typically, enquirers are physicians posing patient-specific queries, requiring the staff members to search for evidence, read and interpret the scientific literature on the topic, and produce a verbal and/or written response to the query. Motives for applying for such information may be the need for evidence-based information and decision support, and the lack of time or skills to search the literature.[6]

Seven DICs chose to participate in the study. During the 8-week study period (January–March 2013), the staff members reported the estimated time consumption per response to all queries. The time consumed per query included time needed to search for, read, and interpret the literature; writing a response; quality assurance by co-signature or discussions with a colleague and all administrative tasks (eg, indexing the query and response in a database); and communication with the enquirer. The study was reported to Norwegian Social Science Data Services in accordance with national legislation.

Members of the project group delineated six fictitious study queries (Table I), all presented in the Norwegian, Swedish, and Danish languages. During the study period, GPs familiar with each center simultaneously posed these study queries by e-mail to all centers. We chose to pose a total of six queries, as we did not want to give a large amount of extra work to the centers. Staff members at the participating centers were, for ethical reasons, informed about the study, but were blinded in terms of how many and which queries were study queries and how they were going to be assessed. By posing the same study queries to all centers, we were able to compare the responses processed by the different centers.

Staff members at all seven centers produced responses to each of the six queries, and the GPs received written responses by e-mail. The 42 responses were distributed to the project leader (L.A.R.), who anonymized the responses for staff members and centers and distributed them to seven internal experts and six external experts who were recruited via the leaders of the centers and who were familiar with the services, for the assessment of quality. For the four centers in Norway, this meant that the experts reviewing them would be completely blinded by which center had produced which answer. For the other three centers, it was not possible to anonymize them completely, as only one center in Sweden participated, giving their answers in Swedish, and the two Danish centers had very different styles, one of them giving the main answer in English, the other one in Danish.

All 13 experts individually assessed the quality of the responses, using a survey form with detailed quality criteria. We developed the survey based on our knowledge and experience from the DICs query-answer service. The specific quality criteria are given in Table II. Most criteria were scored on a five-point scale ("in a very small degree" to "to a very large degree"). In retrospect, all responses were converted to numbers 0 = poorest quality to 4 = best quality.

Statistical Analysis

We performed mixed-model linear regression analyses with total time as the independent variable, and the common quality score, external experts' sum score, internal experts' sum score (Table II), and overall impression of the response, respectively, as dependent variables. These analyses included crossed random effects of centers, queries, and individual experts. A statistical interaction between internal/external experts and total time consumption was included when relevant. These analyses used restricted maximum likelihood estimation. We performed the analyses with and without the inclusion of the center with the highest mean time of consumption, and with and without controlling for the staff members' experience (<2 or ≥2 years).

To quantify the extent of agreement among experts, we used the Kendall's coefficient of concordance (W) and Gwet's agreement coefficient with first-order chance correction (AC1). Kendall's W was chosen because we present ordinal variables assessed by multiple experts.[21] It quantifies the extent of agreement with respect to an expert's ranking of subjects,[22] in this case, responses. It becomes, however, increasingly difficult to achieve high W scores as the numbers of experts increase.[21] For the dichotomous variables, we report both the absolute agreement and Gwet's AC1. The absolute agreement

Table I. The six drug-related queries from general practitioners (GPs) were simultaneously submitted to seven Scandinavian drug information centers (DICs) during a study period of 8 weeks.[*]

Query No.	Query	Category
I	A female patient presents with deep, infected pockets close to the jaw bone, and needs to have these rinsed every 4th to 5th week for the next 6 months. The patient uses alendronate 70 mg once weekly. Should alendronate be discontinued during treatment?	Adverse effects; pharmacokinetics
II	A male patient with formerly performed gastric bypass needs treatment for *Helicobacter pylori* infection due to symptomatic peptic ulcer. The GP wants to start treatment with pantoprazole 40 mg once daily, metronidazole 400 mg BID, and amoxicillin 750 mg BID. Will the absorption of the drugs be reduced, and are dosage adjustments necessary?	Pharmacokinetics
III	A pregnant woman manifests with moderate depression (MADRS score 29), and there is indication for treatment with an antidepressant. What antidepressant is the first choice of drug during pregnancy?	Pregnancy
IV	A GP has registered an increasing use of *Ginkgo biloba* in nursing care homes and home nursing services. He (she) has also registered that ginkgo might increase bleeding time. What documentation exists on this topic, and what is the relevance for concurrent use of, for example, warfarin, acetylsalicylic acid, clopidogrel, and enoxaparin?	Herbal medicines; adverse effects; drug interactions
V	A male patient, aged 75 years, has gradually developed impaired cognition for the last 5–6 months (MMS score of 18 at examination). He has essential hypertension treated with atenolol 100 mg once daily and losartan/hydrochlorothiazide 100/12.5 mg once daily. His blood pressure was 130/90 mm Hg at the latest appointment. He also uses simvastatin 40 mg and acetylsalicylic acid 160 mg once daily. He uses paroxetine 40 mg in the morning for anxiety/depression, diazepam 5 mg as needed and promethazine/propiomazine[†] 50 mg at night for sleep. He also uses tolterodine 2.8/4 mg[‡] once daily for overactive bladder (dosage increased from 1.4/2 mg[‡] 3 months ago). The patient does not smoke. Can any of these drugs, or drug interactions, increase the risk for impaired cognition?	Drug use in the elderly; drug interactions; adverse effects
VI	A female patient, 13 weeks postpartum, presents with active ulcerative colitis. She has earlier been treated with sulfasalazine 500 mg × 3 (discontinued during pregnancy). Can she use sulfasalazine while breast-feeding?	Breast-feeding

MADRS = Montgomery and Asberg Depression Rating Scale; MMS = Mini-Mental status.
[*]All GPs asked each query on the same day. Queries were originally asked in the Norwegian, Danish, and Swedish languages.
[†]Promethazine is marketed in Norway and Denmark, but not in Sweden. For the Swedish query, we used the drug propiomazine. Both drugs are derivatives of phenothiazine with antihistaminic and anticholinergic effects.
[‡]Tolterodine is marketed as 2- and 4-mg depot capsules in Norway and Sweden, and as 1.4- and 2.8-mg depot capsules in Denmark.

Table II. Scores of detailed quality criteria used to assess the quality of responses to six drug-related queries posed to seven Scandinavian drug information centers (DICs).

Criterion No.[*]	Quality Criterion	Internal Expert Evaluation			External Expert Evaluation			Total Kendall's W[†]
		Score, Mean (SD)	n	Kendall's W[†]	Score, Mean (SD)	n	Kendall's W[†]	
1	Overall impression of the response	2.6 (1.0)	294	0.34	2.6 (1.0)	252	0.04	0.20
2	Relevance of the response	3.2 (0.7)	291	0.57	3.1 (0.8)	252	0.35	0.44
3	Conclusions/advice are in accordance with the presented documentation	2.9 (0.8)	289	0.33	2.8 (0.8)	251	0.15	0.25
4	The extent of the response	3.1 (1.1)	293	0.07	3.2 (1.2)	252	0.17	0.11
5	Unnecessary information in the response	3.0 (0.8)	291	0.21	3.0 (1.0)	251	0.22	0.21
6	Readability of the response	2.8 (0.7)	294	0.23	2.7 (0.9)	251	0.22	0.21
7	Logical structure of the response	2.8 (0.7)	293	0.31	2.8 (0.8)	252	0.17	0.23
8	Clinical usefulness of the response	NA	–	NA	2.9 (0.9)	252	0.20	–
9	Would you feel safe making a decision concerning the case, based on the response?	NA	–	NA	2.6 (1.1)	247	0.12	–
10	Understanding of the response	NA	–	NA	3.1 (0.7)	251	0.43	–
11	Staff member provided extra value to the response, besides presenting data[‡]	1.9 (1.0)	294	0.26	NA	-	NA	–

(continued)

Table II. (continued).

Criterion No.[*]	Quality Criterion	Internal Expert Evaluation			External Expert Evaluation			Total Kendall's W[†]
		Score, Mean (SD)	n	Kendall's W[†]	Score, Mean (SD)	n	Kendall's W[†]	
12	All relevant aspects of the case covered	2.7 (0.8)	293	0.31	NA	-	NA	-

		Internal Expert Evaluation			External Expert Evaluation			
		No of "Yes" Responses, %		Gwet's AC1[§], (95% CI)	No of "Yes" Responses, %		Gwet's AC1[§], (95% CI)	Total Gwet's AC1[§], (95% CI)
13	Advice on how to handle the case was given	247 (84.0)	294	0.76 (0.63–0.88)[‖]	201 (79.8)	252	0.73 (0.59–0.86)[¶]	0.73 (0.61–0.86)[¶]
14	No incorrect statements in response	256 (88.9)	288	0.81 (0.71–0.92)[#]	NA	-	NA	-
15	No statements that might be misunderstood in response	282 (95.9)	294	0.92 (0.86–0.97)[**]	NA	-	NA	-
16	No expressions or abbreviations that should have been explained in response	277 (94.2)	294	0.88 (0.82–0.94)[††]	NA	-	NA	-

NA = not applicable.
[*]Criteria 1 to 3 and 5 to 12 were scored from 0 (poorest) to 4 (best); criterion 4 was scored 0 (poorest), 2, or 4 (best); and criteria 13 to 16 were scored either 0, meaning no (poorest) or 4, meaning yes (best).
[†]Kendall's coefficient of concordance, for ordinal variables assessed by multiple experts (value, 0–1).
[‡]That is, interpreting data and evaluating literature critically.
[§]Gwet's agreement coefficient with first-order chance correction, AC1 (value, 0–1).
[‖]Absolute agreement between experts (value, 0–1), 0.82.
[¶]Absolute agreement between experts, 0.81.
[#]Absolute agreement between experts, 0.85.
[**]Absolute agreement between experts, 0.93.
[††]Absolute agreement between experts, 0.89.

Table III. Time consumption for seven Scandinavian drug information centers (DICs) answering six specific drug-related queries.

Query No.	Requested Time Frame	Time Consumption per Response (h:min)				Mean Response Time (d)
		Mean	Median	Minimum	Maximum	
I	Within a week	05:16	03:35	01:35	12:36	3.8
II	Within 2 days	04:29	03:44	02:00	10:39	1.4
III	Within the next day	01:38	01:09	00:17	03:30	0.1
IV	None	03:59	02:05	00:42	16:04	4.0
V	None	04:50	04:08	02:06	12:25	4.1
VI	None	06:38	02:50	00:35	24:17	1.4
All queries	–	04:26	03:08	00:17	24:17	2.6

value provides a sensible value of unity when all experts agree; however, absolute agreement does not adjust for agreement by chance.[23]

Of 42 cases, 2 were excluded from the regression analysis due to lack of time estimates. In 2 other cases, the time component for quality assurance was missing, and we imputed this component based on all of the other time components of all of the other cases. For detailed quality scores, 26 single values were missing of 7800 possible values, representing 0.3% missing values. The missing values represented 21 of 520 individual expert evaluations (4%). We used the expectation-maximization algorithm to do single imputation. The quality scores of the internal experts were used for predicting the scores of the missing values from the internal experts, and the quality scores of the external experts were used for predicting the scores of the missing values from the external experts. Two imputed values were higher than the permissible range, and these were set to the highest possible value (ie, 4). For the calculations of agreement, all imputed values were rounded off to the nearest integer represented in the score of the actual quality criteria.

Statistical analyses were performed using SPSS version 21 (IBM Corp., Armonk, New York), STATA IC version 13 (StataCorp LP, College Station, Texas), and AgreeStat 2013.3 (Advanced Analytics, LLC, Gaithersburg, Maryland). P values <0.05 were considered statistically significant.

RESULTS

Both expert groups assessed the quality of the responses as "satisfactory" to "good." The seven responses to each of the six queries were generally professionally concordant, although some exceptions did occur. The interpretation of the literature, assessment of clinical significance of the therapeutic problems, and management recommendations varied between responses to the same query, as did both the internal and external experts' interpretation of these issues.

Table II shows the results of the detailed quality criteria as well as the measures of concordance between experts. Generally, the scores given on each of the quality criteria varied more among the external than among the internal experts. Of the 42 responses, 27 (64%) were quality-assured using co-signatures, and 24 (57%), by discussions with colleagues. In 16 responses (38%), both methods were used. Five responses (12%) were not quality-assured by others, whereas information was missing in two cases (5%). Of the four responses with lowest quality scores, none were quality-assured using co-signature, but three had been discussed with colleagues.

Responses to the queries were given within a mean of 2.5 days, with a minimum of 0 days (ie, the answer was given on the same day) to a maximum 12 days. For the three queries in which a response was requested within a certain time frame, all responses except one were delivered within the requested time. Mean, median, and minimum and maximum time consumptions, requested time frame and response time for each query are shown in Table III. Four centers spent a mean of 2 to 3 hours per query, two centers spent a mean of 3 to 4 hours per query, and one center spent a mean of >13 hours per query.

Table IV. Summarized quality scores for internal (clinical pharmacologists) and external (general practitioners) experts reviewing responses to six drug-related queries posed to seven Scandinavian drug information centers (DICs).[*]

Query No.	Internal Experts' Sum Score[†]			External Experts' Sum Score[‡]		
	Mean (%)	Minimum (%)	Maximum (%)	Mean (%)	Minimum (%)	Maximum (%)
I	37.0 (78)	33.3 (69)	40.9 (85)	26.7 (67)	21.3 (53)	33.0 (83)
II	37.0 (78)	28.2 (59)	41.2 (86)	29.9 (75)	17.2 (43)	35.2 (88)
III	38.1 (79)	35.4 (74)	40.8 (85)	32.0 (80)	27.8 (70)	36.2 (91)
IV	36.2 (75)	30.8 (64)	40.3 (84)	27.7 (69)	20.3 (51)	30.8 (77)
V	35.8 (75)	30.0 (63)	39.6 (83)	28.4 (71)	22.0 (55)	35.0 (88)
VI	38.7 (81)	35.1 (73)	40.3 (84)	32.5 (81)	27.2 (68)	36.8 (92)
All queries	37.1 (77)	28.2 (59)	41.2 (86)	29.5 (74)	17.2 (43)	36.8 (92)

[*]The possible range of sum score was 0 to 40 for the external experts' score and 0 to 48 for the internal experts' score. The percentages represent the specific score in percentage of the maximum possible score in the same category.
[†]Criteria 2 to 7 and 11 to 16 (Table II) were included in the internal experts sum score.
[‡]Criteria 2 to 10 and 13 (Table II) were included in the external experts sum score.

Table IV shows the results of the summarization of the internal experts' quality score and external experts' quality score, respectively. Table V shows the results of the mixed model linear regression analyses of time as a predictor variable for various quality measurements. Sensitivity analysis excluding the center with the highest time consumption did not change the results significantly, except that the external experts' score were statistically significantly associated with a decreasing time consumption. The inclusion of staff members' experience did not change the results (data not shown).

Table V. Results from mixed model linear regression analyses exploring whether time consumption predicts different measures of quality of written answers to six drug-related queries posed to seven Scandinavian drug information centers (DICs) (n = 40; data from 2 cases are unavailable).[*]

Dependent Variable	Possible Score	Estimated Change in Quality Score per Extra Hour Spent on Response, (95% CI)	P
Internal experts' common quality score[†]	0–28	–0.04 (–0.17 to 0.09)	0.52
External experts' common quality score[†]	0–28	–0.20 (–0.33 to –0.06)	0.004
Internal experts' sum score[‡]	0–48	–0.06 (–0.24 to 0.13)	0.56
External experts' sum score[§]	0–40	–0.21 (–0.47 to 0.04)	0.10
Internal experts' overall impression of response	0–4	0.01 (–0.02 to 0.04)	0.47
External experts' overall impression of response	0–4	–0.05 (–0.08 to –0.01)	0.005

[*]Sums of quality scores were calculated using different combinations of 16 individual quality criteria assessed by internal experts (clinical pharmacologists) and external experts (general practitioners).
[†]Criteria 2 to 7 and 13 (Table II) was included in the score.
[‡]Criteria 2 to 7 and 11 to 16 (Table II) was included in the score.
[§]Criteria 2 to 10 and 13 (Table II) was included in the score.

DISCUSSION

To our knowledge, this is the first study to explore the association between time consumption when processing responses to drug-related queries to DICs, and the quality of the responses. The results are somewhat surprising; we found no association between time consumption and quality as assessed by the internal experts. For the external expert quality evaluation, however, we found that an increase in time consumption was statistically significantly associated with a *decrease* in the assessed quality, both when measured as the common quality score and as the overall impression of the response. The clinical importance of these findings, however, is probably negligible. The external experts' common quality score was reduced with 0.2 per extra hour of work, which means that working an extra 5 hours with a response would decrease the score with 1.0 (of a possible maximum of 28.0), a change not considered clinically important. Certainly, these results do not mean that the longer the time one spends processing a query, the poorer the response. In addition, we cannot extrapolate these results outside the time intervals reported in this study. We postulate that during the first minutes or hours of work, the quality increases in parallel with the time consumption, but at some point additional time invested may not further improve quality. Moreover, this finding could be of particular relevance for the external experts. For these experts, there was a tendency that the assessment of readability, usefulness, and understanding of the response decreased with increasing time consumption, as did the assessment of whether they would feel safe making a decision concerning the case. The responses with the highest time consumption might be more extensive and contain more details on, for example, study results, which might make them more difficult to read. The length (number of words) of the responses certainly increased with time consumption. Exploring the association between the length of the responses and the assessed quality, however, we found no trend that the length of the responses affected the assessment of the quality of the responses, among either internal or external experts. It should however be remembered that our sample size was small, and a lot of the total variance in quality scores could not be accounted for in our analyses.

Compared with previous studies that used fictitious queries to assess the quality of the DIC services,[15–19] the overall quality of the responses in the present study was satisfactory. All centers produced responses to all study queries, and the mean scores were mostly satisfactory to good. The response times, a factor potentially important for the usefulness of such responses,[6] were generally satisfactory, and with one exception, all queries were processed within the requested timeframe. Adding the response time to the multiple regression analyses for the external experts (who were informed of this) had no effect on the association between time consumption and the common quality score, or the external experts' sum score. However, the association between time consumption and total impression of the response did not reach statistical significance when adjusted for response time (results not shown). Although the external experts were not actual users of the services, they may have emphasized the response time, particularly in the assessment of the total impression.

Former studies of the US DICs, although quite old, have assessed the correctness of responses.[17–19] We did not assess the detailed correctness of the responses, as the queries were complex and had no clear-cut answers. We chose this approach because queries demanding simple statements of facts are not representative of those normally posed to the DICs in Scandinavian countries. Therefore, we also did not produce a "correct" control response. We did ask the internal experts whether they identified incorrect statements in the responses, and such statements were reported in 32 of 288 ratings (11%). Different experts, however, actually disagreed on a few statements in terms of correctness. For example, the clinical relevance of a pharmacokinetic drug interaction between paroxetine and atenolol was deemed to be clinically significant by some experts, who rewarded the responses that made this interpretation, whereas others assessed this drug interaction to be clinically nonsignificant, discrediting these responses. This example reflects how challenging it is both to respond to these queries and to assess the responses. It is also complicated to make objective quality criteria for DICs' responses to complex queries, as the assessment of quality is clearly individual and dependent on the experts' knowledge, experience, and attitudes.

By comparing different responses to the same queries, we would theoretically be able to determine which elements increase or decrease the quality of responses. The responses with the lowest quality were

quite easy to identify, whereas high-quality responses were more difficult to point out, as the differences between scores of individual responses were small. Comparing high-quality to low-quality responses, the quality criteria with the largest difference in numeric score was whether advice was given, and the degree to which advice was given certainly varied between responses. The differences were also large in terms of whether external experts would feel safe making a decision concerning the cases.

Of the four responses that, based on summarized scores from both the internal and external experts, had the lowest quality, none were quality-assured using co-signature, but three were discussed with colleagues. Although the number of responses is small, this small number may suggest that the use of co-signatures is a preferable way of quality-assuring the written responses. The use of co-signature implies that another colleague has read the original query and the suggested written response and discussed these with the staff member primarily involved in the query. The results also imply that oral discussions with colleagues may not always be sufficient for quality assurance, which is especially true for the centers that primarily give written responses; putting your signature on a paper might to a larger extent involve an obligation or a commitment.

The agreement within both internal and external experts was generally low in this study. This finding is not surprising, as the assessment of quality of drug information responses is very subjective.[8,24] Looking at the results of the dichotomous variables, there was a higher level of agreement. For future studies, however, adding a more formal qualitative approach may be a valuable supplement as it allows for more thorough follow-up of experts' assessments and comments.

Time consumption when answering the study queries varied considerably, as did the time consumption in responding to "real" queries, reported in our previous study.[20] The mean response time was higher for these study queries than for the real queries processed during the same time period (266 and 178 minutes, respectively). The reasons for this difference might be that the center with the clearly highest time consumption had a greater proportion of the responses in the present study. Also, a large proportion of the study queries (22 [52%]) was handled by staff members with 0 to 2 years of experience in the present study, a factor shown to predict an increase in time consumption.[20]

Study Strengths and Limitations

We used an experimental, prospective design as suggested by Spinewine et al,[13] giving us extensive control over the collected data. We ascertained that the collected responses were the ones that were actually given to the enquirers, and the staff members were blinded in terms of which queries were study queries. We have no confirmative knowledge that the blinding was violated. We cannot exclude that staff members have worked in a way different from usual during the study period to increase the quality of the responses, and that this factor may have increased the time consumption. However, we consider the duration of the study period (8 weeks) being a strength, as we believe the staff members got used to both the estimation of time consumption and the thought of being evaluated, which may have decreased the suspiciousness toward queries as the study period went on.

In addition to different DICs' styles and staff members' experiences,[20] one possible explanation for the great variation in time consumption is the staff members' former knowledge of the specific topics. Some staff members may have spent a short time giving a good-quality response because they were familiar with the documentation. We were unable to control for this factor, as we have no registration of each staff member's personal knowledge within the area. Another inaccuracy of the multivariate analyses is that we did not control for the fact that some staff members answered more than one query.

External reviewers like our GPs might help reduce bias[13] and are particularly suitable for written responses.[9] Our external experts were recruited via the leaders of the centers and were familiar with the services. As such, we do not know whether they were completely unbiased, especially toward their own center. The same argument is valid also for the internal experts, who were clinical pharmacologists recruited from five of the participating DICs. We did not adjust for this factor in the multivariate analyses. However, all experts were blinded at least to the responses from the four DICs in Norway.

The two expert groups were intended to review the quality of responses from different angles: The internal experts had special competence in the field of

clinical pharmacology and should be suitable for assessing the correctness and completeness from a scientific point of view. The external experts represented the users of the centers and should be suitable for assessing, for example, the usefulness of the responses. In retrospect, the members of the two groups of experts may have been more similar in their focus than expected. Our external experts may not have been representative of all GPs, as several of them had a PhD and were working part-time in academia. Still, there are differences in the results between these two groups that may reflect their different perspectives.

We have used the word *quality* in its broadest sense, as the "standard of something, measured against other thing of a similar kind."[25] The use of explicit criteria has been discussed by others.[24] One might argue that the most important quality criterion is what actually happens to the patients of health care professionals posing queries to DICs.[2] However, due to the lack of controls and because one response could affect several future patients, it is a challenge to measure the impact of these services in such a perspective. Our quality criteria were based on data from earlier studies of quality assessments of DIC responses,[2,10,15–19] as well as our own knowledge of the services. Our choice of criteria has not been validated, but as pointed out, there is also a lack of widely accepted criteria.[1,13,24] We do not know how the individual experts interpreted the assessment criteria, and clearly, different experts may have emphasized different aspects of the responses. In retrospect, however, we observed that the quality criterion "overall impression of the response" might be just as suitable for measuring the overall quality as any of the summarized scores.

CONCLUSIONS

The quality of responses processed by the DICs in Scandinavian countries was generally satisfactory to good. We found large variances in both style and content, as well as in the time consumed to process them. We found no indication that increasing time consumption also increased the quality of the responses; however, differences in quality were small, and the findings should be interpreted with caution. Other factors that may be more or less associated with an increasing time consumption may be more important for quality (eg, the digestion of documentation, giving specific advice when possible, and the use of co-signatures for quality assurance).

CONFLICTS OF INTEREST
L.A. Reppe received support from the Norwegian RELIS for the submitted work; J. Schjøtt, P. Damkier, H.R. Christensen, J.P. Kampmann, and O. Spigset are employees of the participating DICs. Y. Böttiger was an employee of a participating DIC. All of the authors designed the study, L.A. Reppe collected the data, and L.A. Reppe and S. Lydersen analyzed the data. All of the authors contributed to the writing of the manuscript and to the decision to submit the manuscript for publication. The authors have indicated that they have no other conflicts of interest with regard to the content of this article.

REFERENCES
1. American Society of Health-System Pharmacists (ASHP). ASHP guidelines on the pharmacist's role in providing drug information 2014. http://www.ashp.org/DocLibrary/BestPractices/SpecificGdlMedInfo.aspx. Accessed June 21, 2016.
2. Repchinsky CA, Masuhara EJ. Quality assurance program for a drug information center. *Drug Intell Clin Pharm*. 1987;21:816–820.
3. Schjøtt J, Reppe LA, Roland PD, et al. A question-answer pair (QAP) database integrated with websites to answer complex questions submitted to the Regional Medicines Information and Pharmacovigilance Centres in Norway (RELIS): a descriptive study. *BMJ Open*. March 15, 2012;2: e000642. [Epub ahead of print].
4. Frost Widnes SK, Schjøtt J. Drug use in pregnancy– physicians' evaluation of quality and clinical impact of

drug information centres. *Eur J Clin Pharmacol*. 2009;65:303–308.

5. Schjøtt J, Pomp E, Gedde-Dahl A. Quality and impact of problem-oriented drug information: a method to change clinical practice among physicians? *Eur J Clin Pharmacol*. 2002;57:897–902.

6. Hedegaard U, Damkier P. Problem-oriented drug information: physicians' expectations and impact on clinical practice. *Eur J Clin Pharmacol*. 2009;65:515–522.

7. McEntee JE, Henderson SL, Rutter PM, et al. Utility and value of a medicines information service provided by pharmacists: a survey of health professionals. *Int J Pharm Pract*. 2010;18:353–361.

8. Johnson N, Dupuis LL. A quality assurance audit of a drug information service. *Can J Hosp Pharm*. 1989;42:57–61.

9. Thompson DF, Heflin NR. Quality assurance in drug information and poison centers: a review. *Hosp Pharm*. 1985;20:758–760.

10. Bertsche T, Hammerlein A, Schulz M. German national drug information service: user satisfaction and potential positive patient outcomes. *Pharmacy World Sci*. 2007;29:167–172.

11. Bramley D, Mohandas C, Soor S, et al. Does a medicines information service have a positive impact on patient care? *Pharm J*. 2009;282:139–140.

12. Innes AJ, Bramley DM, Wills S. The impact of UK medicines information services on patient care, clinical outcomes and medicines safety: An evaluation of healthcare professionals' opinions. *Eur J Hosp Pharm*. 2014;21:222–228.

13. Spinewine A, Dean B. Measuring the impact of medicines information services on patient care: methodological considerations. *Pharmacy World Sci*. 2002;24:177–181.

14. Lyrvall PH. Report on a questionnaire study among users of the Drug Information Centre in Stockholm, January to July 1990, performed by the Drug Information Centre and the Institute of Clinical Pharmacology, Huddinge Hospital. In: Problem-oriented drug information – an opportunity to improve the quality of hospital care. Huddinge University Hospital; 1994.

15. Gallo GR, Vander Zanden JA, Wertheimer AI. Anonymous peer review of answers received from drug information centres. *J Clin Hosp Pharm*. 1985;10:397–401.

16. Del Mar CB, Silagy CA, Glasziou PP, et al. Feasibility of an evidence-based literature search service for general practitioners. *Med J Aust*. 2001;175:134–137.

17. Calis KA, Anderson DW, Auth DA, et al. Quality of pharmacotherapy consultations provided by drug information centers in the United States. *Pharmacotherapy*. 2000;20:830–836.

18. Beaird SL, Coley RM, Blunt JR. Assessing the accuracy of drug information responses from drug information centers. *Ann Pharmacother*. 1994;28:707–711.

19. Halbert MR, Kelly WN, Miller DE. Drug information centers: lack of generic equivalence. *Drug Intell Clin Pharm*. 1977;11:728–735.

20. Reppe LA, Spigset O, Böttiger Y, et al. Factors associated with time consumption when answering drug-related queries to Scandinavian drug information centres: a multi-centre study. *Eur J Clin Pharmacol*. 2014;70:1395–1401.

21. Gisev N, Bell JS, Chen TF. Inter-rater agreement and interrater reliability: key concepts, approaches, and applications. *Res Social Adm Pharm*. 2013;9:330–338.

22. Gwet KL. Handbook of Inter-rater reliability. The definitive guide to measuring the extent of agreement among raters, 4th ed. Gaithersburg, Md: Advanced Analytics; 2014.

23. Andreasen S, Backe B, Lydersen S, et al. The consistency of experts evaluation of obstetrics claims for compensation. *BJOG*. 2015;122:948–953.

24. Wheeler-Usher DH, Hermann FF, Wanke LA. Problems encountered in using written criteria to assess drug information responses. *Am J Hosp Pharm*. 1990;47:795–797.

25. Oxford University Press. Quality. Oxford Dictionaries http://www.oxforddictionaries.com/definition/english/quality?searchDictCode=all2014. Accessed June 21, 2016.

Address correspondence to: Linda Amundstuen Reppe MSc.Pharm, Nord University Faculty of Health Sciences, Post Box 2501, Steinkjer 7729, Norway. E-mail: linda.a.reppe@nord.no