Inter-rater reliability of assessments regarding the quality of drug treatment, and drug-related hospital admissions

Johan Lönnbro\textsuperscript{1,2} | Lina Holmqvist\textsuperscript{2,3} | Elisabeth Persson\textsuperscript{2} | Per Thysell\textsuperscript{2} | N. David Åberg\textsuperscript{2,3} | Susanna M. Wallerstedt\textsuperscript{1,4}

\textsuperscript{1}Department of Pharmacology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
\textsuperscript{2}Department of Acute Medicine and Geriatrics, Region Västra Götaland, Sahlgrenska University Hospital, Gothenburg, Sweden
\textsuperscript{3}Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska University Hospital, Gothenburg, Sweden
\textsuperscript{4}HTA Centre, Sahlgrenska University Hospital, Gothenburg, Sweden

Correspondence
Johan Lönnbro, Department of Pharmacology, Sahlgrenska Academy, University of Gothenburg, SE-405 30, Gothenburg, Sweden. Email: johan.lonnbro@gu.se

Aims: To investigate inter-rater agreement on the quality of drug treatment, and the relationship between the drug treatment and hospital admission.

Methods: Three specialist physicians and two resident physicians determined, independently and in consensus, the quality of drug treatment from an overall medical perspective, and its association with admission, in 30 randomly selected patients (50\% female, median age 72 years) admitted to Sahlgrenska University Hospital, Sweden, in April 2018. The inter-rater agreement was evaluated with Gwet’s agreement coefficient (AC\textsubscript{1}).

Results: In all, 200 (95\%) out of 210 drugs at admission and 238 (97\%) out of 245 drugs at discharge were assessed as reasonable drug treatment by all assessors. Conversely, none of the drugs at admission, and two at discharge, were assessed as unreasonable drug treatment by all assessors (AC\textsubscript{1}: 0.88 and 0.94 [all], 0.86 and 0.95 [specialists], 0.92 and 0.92 [residents], respectively). The assessments regarding the association between the drug treatment and the hospital admission (not related or main/contributory reason) were consistent between the assessors for 16 out of 30 patients (AC\textsubscript{1}: 0.67 [all], 0.74 [specialists], 0.54 [residents]). In none of the three cases where the hospital admission was considered possibly attributable to a prescribing error did the assessors make consistent assessments.

Conclusions: As the inter-rater agreement ranged between weak and almost perfect, the reliability of assessments of drug treatment quality, as well as adverse consequences, appears to be a methodological concern. To yield acceptably reliable results regarding both drug treatment aspects at issue, specialist physicians should be involved.

Keywords: adverse drug reaction, inter-rater agreement, pharmacotherapy, prescribing quality
1 | INTRODUCTION

Prescribing of medicines involves major challenges and in an endeavour to improve prescribing practices, prescribing quality and adverse consequences have been extensively studied. However, there is no clear-cut distinction between correct and incorrect drug treatment, and the association between a drug and an adverse reaction is rarely certain. Inherent to the nature of pharmacotherapy, subjective assessments play an important role in the scientific literature.

The basis for determining the quality of drug treatment, and consequences of prescribing, is a medical assessment, taking into account individual aspects such as comorbidities, concurrent drug treatment and treatment effects. Indeed, evidence-based guidelines often match poorly to complex multimorbidity, and the ability of easy-to-apply indicators of prescribing quality, identifying potentially inappropriate medications (PIMs) and potential prescribing omissions (PPOs), to reflect quality of drug treatment is limited. In addition, several treatment alternatives may be reasonable at the individual level. Further, adverse drug reactions (ADRs) are often difficult to distinguish from other medical conditions, including spontaneous emergence and worsening of present diseases. Indeed, in most ADR reports, the association between the drug and the reaction is categorised as “possible”, which, according to the nomenclature used by the World Health Organization (WHO), implies that a disease may just as well explain the course of events.

The scientific literature regarding the extent of subjectivity involved in assessments of quality of drug treatment and drug-related adverse events, i.e., the inter-rater reliability, is surprisingly limited. Although evidence regarding reliability has been provided for the application of PIM/PPO tools, there have, to the best of our knowledge, been no scientific publications focusing on the assessment of the quality of the entire drug treatment, from an overall medical perspective. Regarding assessments of the prevalence of ADRs and drug-related admissions, some studies report kappa statistics. Consistently, however, the number of raters has been low, often restricted to two, while in studies with multiple assessors the study population has been restricted from the start. Indeed, the assessments in the four studies just cited were performed in selected patients who had already been determined to present with potential drug-related issues. As far as we are aware, studies focusing on the inter-rater reliability between multiple assessors with different levels of medical experience, assessing unselected patients in a broad hospital setting, are lacking.

To fill this gap, we conducted the present study in a random sample of hospital admissions, with the aim to evaluate inter-rater agreement on the quality of drug treatment as well as the relation between the drug treatment and the hospital admission, including preventability.

2 | METHODS

The patients (n = 30) were randomly selected from a total of 864 individuals, ≥18 years of age, with an unplanned admission to the Sahlgrenska University Hospital over a 2-week period in April 2018. Five physicians independently assessed the quality of each patient’s drug treatment at admission and discharge, as well as the association, if any, between the drug treatment and the hospital admission. After the initial independent assessments, all five physicians together made a consensus assessment. The study procedures are outlined in Figure 1.

2.1 | Data source and extraction

One author (J.L.) extracted all patient data from the electronic health records, including patient characteristics and information regarding the present admission. The data extractions were checked by another author (L.H.). One author (J.L) also recorded PIMs/PPOs according to the Screening Tool of Older Persons’ Prescriptions (STOPP) and Screening Tool to Alert to Right Treatment (START) criteria.
2.2 Assessments

Three internal medicine physicians, one of whom was a senior consultant (N.D.Å.) and two who were specialists (J.L., L.H.), as well as two residents in internal medicine (E.P., P.T.), independently assessed the quality of each patient’s drug treatment at admission and discharge, dichotomised into reasonable and unreasonable. The assessments were made from an overall medical perspective, taking into account the health condition of each patient. The quality assessments at admission, as opposed to discharge, were based solely on information available before the current visit to the hospital. Taking into account additional information revealed during the hospital stay, the physicians also independently assessed whether the drug treatment could have contributed to, or was the main reason for, the admission.27 If so, they decided whether a prescribing error definitely or possibly could have contributed to the admission.28 The physicians had 2.5–19 years’ professional experience of patient work. Despite differences in experience, all were well experienced regarding medical assessments in patient care and the prescribing of drugs. However, none had previous experience of performing the present assessments in a research setting.

The assessors based their assessments on the data extractions described above, including the recorded PIMs and PPOs according to STOPP/START, and information available in the unfiltered electronic health record. To systemise the assessments and to avoid overlooking drug treatment generally considered suboptimal, each assessor independently applied a comprehensive set of indicators of prescribing quality prior to their decision making.29 This tool, provided by the Swedish National Board of Health and Welfare, included 77 PIMs as well as 20 PPOs applicable at the individual level. Although this tool, as well as the STOPP/START criteria mentioned above, are intended for older people, we considered them appropriate to ascertain that general aspects of importance in prescribing were not overlooked, for instance renal function, potential drug interactions, specific drugs of particular concern, and guidelines regarding pharmacotherapy in common conditions. To further systemize the assessments and to identify ADRs related to the hospital admission, an algorithm was applied27 and causality assessed according to WHO criteria.17 For each ADR, where a possible, probable or certain causal relationship between the drug and the reaction was suspected, the assessors independently determined whether the reaction was (i) the main cause or (ii) a contributory reason for, or (iii) not related to, the admission.27,28 The Hallas criteria of preventability were used to assess whether ADR-related admissions were attributable to a prescribing error.30 and if so, this assessment was dichotomised into definitely or possibly.28 Medical conditions without pharmacological treatment were assessed in a similar manner to determine causality, association with the hospital admission, and to

---

**FIGURE 1** Flowchart of the assessments performed in the study* Screening tools applied before the assessments: (i) the STOPP/START criteria were applied by one assessor and identified PIMs/PPOs were shared with the other assessors, and (ii) the indicators of prescribing quality by the Swedish National Board of Health and Welfare were applied by all assessors individually before their assessments. The assessment of prescribing quality was not restricted to PIMs/PPOs identified with the screening tools. PIM, potentially inappropriate medication; PPO, potential prescribing omission; START, Screening Tool to Alert to Right Treatment; STOPP, Screening Tool of Older Persons’ Prescriptions; WHO-UMC, WHO-Uppsala Monitoring Centre.
what extent a prescribing error, in that case a prescribing omission, could be considered the underlying cause.

In the final consensus discussion, after all the individual assessments had been performed, the five assessors agreed on the quality of drug treatment at admission and discharge, as well as on whether the drug treatment was the main cause or a contributory cause of the admission, and, where an association between the drug treatment and admission was found, to what extent the admission was attributable to a prescribing error.

2.3 | Statistics

Statistical analyses were performed with SPSS Statistics for Windows, version 26–27 (IBM SPSS, Armonk, NY, USA), and in R (R Core Team, 2020). Gwet’s agreement coefficients (AC1), with corresponding 95% confidence intervals (CIs), were calculated to quantify the inter-rater reliability of the assessments.31

Gwet’s AC1 is a chance-corrected method for assessing agreement, similar to Cohen’s kappa. We chose Gwet’s AC1 as the trait prevalence was skewed, with a large proportion of the ratings falling into one of the categories. In such cases, Cohen’s kappa generates low kappa coefficients despite obvious high agreement between the raters, while Gwet’s AC1 is more resistant to this phenomenon.31

To estimate the agreement between groups of assessors (all/specialists/residents) and the consensus evaluations, we used the average of all individually calculated AC1 coefficients between assessors and consensus within the groups. The inter-rater agreement, according to Gwet’s AC1, was interpreted as none (<0.20), minimal (0.21–0.39), weak (0.40–0.59), moderate (0.60–0.79), strong (0.80–0.89) and almost perfect (>0.90).32 We used the R package “Power3Cats” to calculate the sample size required to study inter-rater agreement.33 To have >80% power to detect a statistical difference between kappa values of 0.40 and 0.70, with an error margin of 5%, five raters and three mutually exclusive categories, at least 25 patients had to be assessed. Sample size estimations based on kappa have previously been used for Gwet’s AC1.34

3 | RESULTS

Characteristics of the patients are presented in Table 1. In brief, 50% were women and the median age was 72 years, 19 patients being 65 years or older. The median number of drugs taken regularly at the time of admission was 7. Infectious diseases were the most common direct reason for admission (n = 11). The most common comorbidities were hypertension (n = 14), cardiovascular diseases, including atrial fibrillation, heart failure and ischaemic heart disease (n = 16), diabetes type 2 (n = 6) and active malignancies (n = 6).

In all, 200 (95%) out of 210 drugs at admission, and 238 (97%) out of 245 drugs at discharge, were assessed as reasonable drug treatment by all assessors. Conversely, none of the drugs at admission, and two (0.8%) of the drugs at discharge, were assessed as unreasonable drug treatment by all five assessors. The assessment whether a drug treatment was reasonable or unreasonable for a specific patient differed between the assessors for nine drugs at admission and five drugs at discharge (Table 2). Assessments of drug treatments assessed as unreasonable, common drugs, and drugs categorised differently by the assessors are presented in Table 3.

The drug treatment was consistently considered reasonable by all assessors in 23 patients at admission, and 25 patients at discharge. Conversely, the drug treatment of no patients at admission, and one single patient with two unreasonable medications at discharge, was consistently assessed as unreasonable. The AC1 for assessment of the drug treatment quality at admission and discharge was 0.88 and 0.94, respectively (Table 2). The pairwise AC1 for the group of specialists, regarding treatment quality, ranged between 0.83 and 0.88 at admission and between 0.92 and 0.96 at discharge.

Regarding the relationship between drug treatment and hospital admission, all five assessors made consistent assessments in 16 of the 30 patients. In two of these patients, the drug treatment was considered to have been either the main or a contributory reason for the admission (Table 4)

Both were cancer patients undergoing chemotherapy. The first one was admitted because of neutropenic fever, while the second one, also treated with metoprolol, irbesartan and oxycodone, was admitted because of hypotension, nausea and pain. In the remaining 14 cases, the admission was consistently assessed as not related to the drug treatment. Conversely, in seven of the 14 cases of discordant assessments, the drug treatment was in consensus assessed to be a contributory reason for the admission. The AC1 for the specialist and the resident groups was 0.74 and 0.54, respectively (Table 2). For the residents, the AC1 confidence intervals for assessments of drug treatment quality and the relationship between drug treatment and admission, respectively, did not overlap. For the specialist group, the pairwise AC1, regarding the assessed relationship between drug treatment and the admission, ranged between 0.69 and 0.78. In nine cases, the drug treatment was, at least to some extent, related to the admission, according to the consensus decision (Table 4). The AC1 reflecting the agreement between the specialists’ assessments and consensus was 0.78. The corresponding coefficient for residents was 0.79.

According to the consensus assessment, no hospital admission was definitely attributable to a prescribing error, and the individual assessors did not make consistent assessments in any of the three cases where the consensus assessment considered the hospital admission to be possibly attributable to a prescribing error (Table 4). The AC1 for the specialist and the resident groups was 0.76 and 0.82, respectively (Table 2). Table 4 summarises the reason for admission, the suspected drug treatment, and the individual as well as the consensus assessments in cases where the drug treatment was determined to be related to the admission.
4 | DISCUSSION

Our results show that inter-rater agreement is a non-negligible issue in scientific evaluations of prescribing quality and adverse consequences. With between-assessor AC₁ ranging from 0.54 to 0.95 for various aspects reflecting prescribing practices, the inter-rater agreement ranged between weak and almost perfect. Regarding quality-related prescribing aspects, the inter-rater agreement in the groups of specialists and residents was similar and at an excellent level. Regarding assessments of the association between drug treatment and hospital admissions, the overall agreement was moderate, with moderate and weak agreement between specialists and residents, respectively. The inter-rater agreement between residents was lower for assessments regarding the relationship between drug treatment and admission than for drug treatment quality issues.

The results illustrate that evaluation of drug treatment quality and consequences is a complex issue that requires thorough methodological considerations. Indeed, prevalence figures in descriptive studies on pharmacotherapeutic aspects may be highly dependent on the specific assessor(s). Our findings suggest that assessments of drug treatment quality may be less problematic than assessments of ADRs and their relationship with hospital admissions. With a strong to almost perfect inter-rater agreement and an almost perfect agreement with consensus assessments, it appears that two physicians with at least 2 years of patient work experience, either specialists or residents, may reliably capture quality-related aspects when their overall medical assessment is preceded by the application of screening tools. Assessments of the association between drug treatment and admission, on the other hand, may require more extensive medical competence. With a weak and moderate inter-rater agreement for residents

### TABLE 1 Characteristics of assessed patients (n = 30)

| Characteristic                                      | Median Age (Range), Years | Female Sex, n | Residence/Living Arrangements, n | Median Number of Drugs (Range) | Reason for Admission, n | Admission from the Emergency Department, n | Median Days of Hospital Stay (Range) | Deceased during Hospital Stay, n | Median eGFR at Admission (Range), mL/min/1.73 m² |
|----------------------------------------------------|---------------------------|---------------|---------------------------------|--------------------------------|-------------------------|------------------------------------------|-----------------------------------|---------------------------------|------------------------------------------|
| Age (Range), years                                 | 72 (25-93)                | 15            | Independent Living              | 27                             | Infections             | 24                                        | 3.5 (1-32)                        | 3                               | 78 (12-104)                              |
| Female sex                                         |                           |               | Home Care Service              | 6                              | Cardiovascular Disease |                          |                                    |                                  |                              |
| Residence/Living Arrangements                       |                           |               | Nursing Home                   | 3                              | Abdominal Pain         |                          |                                    |                                  |                              |
| Median Number of Drugs (Range)                      |                           |               | At admission Regular           | 7                              | Tumour-related Symptoms|                          |                                    |                                  |                              |
|                                                   |                           |               | At admission As needed         | 3                              | Trauma                 |                          |                                    |                                  |                              |
|                                                   |                           |               | At discharge Regular           | 6                              | Syncop                |                          |                                    |                                  |                              |
|                                                   |                           |               | At discharge As needed         | 2                              | Other                  |                          |                                    |                                  |                              |
| Reason for Admission                                |                           |               | Hypertension                   | 14                             | Hospital Admission     |                          |                                    |                                  |                              |
|                                                   |                           |               | Atrial Fibrillation            | 9                              | Cardiovascular Disease |                          |                                    |                                  |                              |
|                                                   |                           |               | Heart Failure                  | 4                              | Abdominal Pain         |                          |                                    |                                  |                              |
|                                                   |                           |               | Ischaemic Heart Disease        | 3                              | Tumour-related Symptoms|                          |                                    |                                  |                              |
|                                                   |                           |               | Renal Dysfunction (eGFR<50 mL/min/1.73 m²) | 4                              | Trauma                 |                          |                                    |                                  |                              |
|                                                   |                           |               | Diabetes Type 2                 | 6                              | Syncop                |                          |                                    |                                  |                              |
|                                                   |                           |               | Depression                     | 5                              | Hospital Admission     |                          |                                    |                                  |                              |
|                                                   |                           |               | Sleeping Disorder              | 4                              | Cardiovascular Disease |                          |                                    |                                  |                              |
|                                                   |                           |               | Hypothyroidism                 | 3                              | Abdominal Pain         |                          |                                    |                                  |                              |
|                                                   |                           |               | Active Malignancy              | 6                              | Tumour-related Symptoms|                          |                                    |                                  |                              |

*Severe psoriasis, acute kidney failure, short loss of awareness and spontaneous pneumothorax.

*eGFR, estimated glomerular filtration rate, which was calculated according to the revised Lund–Malmö equation. For one patient, testing of creatinine was not performed during the hospital stay.
### TABLE 2
Assessments of (i) the overall drug treatment quality, and (ii) the relationship between the drug treatment and hospital admission, as well as the inter-rater agreement (AC₁)

| Quality assessment of drug treatment | Assessments (n) | Inter-rater agreement within groups | Agreement versus consensus $^a$ |
|-------------------------------------|-----------------|----------------------------------|-------------------------------|
|                                     | Consensus (n)   | Discordant assessments (n)       | All percent agreement | All AC₁ (95% CI) | Specialists AC₁ (95% CI) | Residents AC₁ (95% CI) | All mean AC₁ | Specialists mean AC₁ | Residents mean AC₁ |
| Admmission                          | Reasonable      | 28 | 5 | 0.90 | 0.88 (0.78–0.98) | 0.86 (0.73–1) | 0.92 (0.81–1) | 0.94 | 0.92 | 0.96 |
|                                     | Unreasonable    | 2  | 2 |      |              |                  |                  |                  |                  |                  |
| Discharge                           | Reasonable      | 28 | 3 | 0.95 | 0.94 (0.87–1) | 0.95 (0.87–1) | 0.92 (0.81–1) | 0.97 | 0.97 | 0.96 |
|                                     | Unreasonable    | 2  | 1 |      |              |                  |                  |                  |                  |                  |
| Hospitalisation                     | Not related to drug treatment | 21 | 7 | 0.75 | 0.67 (0.53–0.82) | 0.74 (0.56–0.92) | 0.54 (0.29–0.80) | 0.80 | 0.78 | 0.79 |
|                                     | Drug treatment main/contributory reason | 1/8 | 0/7 |    |                |                  |                  |                  |                  |                  |
|                                     | Not a prescribing error | 27 | 10 | 0.77 | 0.73 (0.58–0.87) | 0.76 (0.58–0.93) | 0.82 (0.65–0.99) | 0.80 | 0.84 | 0.82 |
|                                     | Definitely/possibly a prescribing error | 0/3 | 0/3 |    |                |                  |                  |                  |                  |                  |

AC₁, Gwet’s agreement coefficient; CI, confidence interval.

$^a$To calculate the inter-rater agreement versus consensus, we used the average of all individually calculated AC₁ coefficients between assessors and consensus within the groups.
and specialists, respectively, and more diverging assessments between residents regarding drug-related admissions compared with drug treatment quality issues despite the use of systematic tools, our results highlight the importance of in-depth medical competence to reliably distinguish adverse reactions from emerging and worsening illness. In fact, the relationship between a drug and a suspected adverse reaction is in most cases categorised as possible, implying that a disease may just as well be the underlying cause. It may not be surprising that the inter-rater agreement was weak in residents. At this stage of their career, although well experienced in patient work and day-to-day pharmacotherapy, residents are still at the beginning of their professional development. As the variation in agreement was limited between all possible combinations of pairs within the group of specialists, the results indicate that two assessors with specialist competence may be sufficient to yield acceptably reliable results regarding drug-related admissions.

Interestingly, for the ten most frequently used drugs, the assessors were in total agreement that the treatment was reasonable in all but four patients, and with these four patients, only one assessor diverged from the rest. This finding suggests a relatively low level of complexity for common drugs. By contrast, the assessors did not make consistent assessments regarding the three drugs that in consensus were determined as unreasonable drug treatment. These drugs were prescribed to two patients, both of whom were being treated with at least eight medications. It may not be surprising that the assessments diverge in such cases; unreasonable drug treatment has been shown to be common among patients with many drugs, which, in turn, has been shown to reflect burden of disease. With multimorbidity and a corresponding extensive medication list, a higher level of complexity can be expected.

The complexity of ADR assessments is also illustrated by the eight cases where the drug treatment was determined in consensus to

---

**TABLE 3** Number of patients on a specific drug for (i) drug treatments assessed in consensus as unreasonable, (ii) drug treatments used in five or more patients, and (iii) drug treatments with at least one discordant assessment regarding reasonableness at either admission or discharge

|                | Admission | Discharge |
|----------------|-----------|-----------|
| Patients on treatment (n) | Unreasonable (consensus) (n) | Discordant assessments (n) | Patients on treatment (n) | Unreasonable (consensus) (n) | Discordant assessments (n) |
| Diazepam       | 1         | Multiple sedatives | 1 | 1 | Multiple sedatives | 0 |
| Alprazolam     | 1         | Multiple sedatives | 1 | 1 | Multiple sedatives | 0 |
| Digoxin        | 1         | Dose too high | 1 | 1 | Dose too high | 1 |
| Paracetamol    | 10        | 0 | 0 | 14 | 0 | 0 |
| Oxycodone      | 8         | 0 | 1 | 9 | 0 | 0 |
| Furosemide     | 7         | 0 | 0 | 8 | 0 | 0 |
| Metoprolol     | 6         | 0 | 1 | 7 | 0 | 0 |
| Zopiclone/Zolpidem | 6/3 | 0/0 | 0/0 | 7/3 | 0/0 | 0/0 |
| Apixaban/Dabigatran | 6/1 | 0/0 | 1/0 | 6/1 | 0/0 | 0/0 |
| Atorvastatin/Simvastatin | 6/2 | 0/0 | 0/0 | 5/2 | 0/0 | 0/0 |
| Cyanocobalamin | 5         | 0 | 0 | 4 | 0 | 0 |
| Omeprazole/Pantoprazole | 3/1 | 0/0 | 0/0 | 6/1 | 0/0 | 0/0 |
| Acetylsalicylic acid | 4 | 0 | 0 | 5 | 0 | 1 |
| Lactulose      | 3         | 0 | 0 | 5 | 0 | 0 |
| Bisoprolol     | 3         | 0 | 0 | 5 | 0 | 0 |
| Oxazepam       | 3         | 0 | 1 | 3 | 0 | 0 |
| Sertraline     | 3         | 0 | 0 | 3 | 0 | 0 |
| Dalteparin/Tinzaparin | 1/0 | 0/0 | 0/0 | 3/1 | 0/0 | 1/0 |
| Amiloride      | 1         | 0 | 1 | 1 | 0 | 0 |
| Methotrexate   | 1         | 0 | 1 | 1 | 0 | 0 |
| Ibuprofen      | 0         | 0 | 0 | 1 | 0 | 1 |

The data in rows 1 to 3 represent the only three cases of unreasonable drug treatment at admission and at discharge (consensus agreement). The data below row 3 are the most frequently used drugs, in descending order of frequency. The last six represent the drug treatments with at least one discordant assessment regarding reasonableness at either admission or discharge.

At admission, discordance between assessors was observed in nine cases, compared with five at discharge.

*For substances belonging to the same pharmacological group, the number of patients on each substance is separated by a slash.
have contributed to the admission, where up to four out of five assessors had a diverging opinion from the start. By contrast, there was no divergence in opinion between assessors regarding the single case where the drug treatment was considered to have been the main cause of admission. This case illustrates the lack of complexity to assess chemotherapy-induced neutropenia, but also the complex weighing of benefits and harms in pharmacotherapy.

In the present study, the inter-rater agreement on potential prescribing errors underlying hospital admissions was comparable with the causality assessments, ranging from moderate to strong. This is consistent with previous studies with multiple assessors, where agreement regarding preventability of adverse drug events was within the same range as agreement regarding causality. Although one study reported lower inter-rater agreement for preventability assessments. In two studies, independent assessments were performed by at least two pharmacists as well as two physicians. The inter-rater agreement on preventability was weak to moderate for pharmacists and minimal to weak for physicians, perhaps illustrating that errors can be pointed out more lightly by those who do not have the prescribing of drugs as their primary professional responsibility.

Our findings indicate that caution is urged in the interpretation of previous research regarding the frequently used outcomes “drug-related admissions” and “prescribing errors”, which have recently been suggested as core outcomes to evaluate interventions for improved prescribing practices. When used, methodological rigour, including the selection as well as the blinding of assessors, will be crucial for reliable results. In fact, our results contribute to the understanding of the largely diverging results regarding the prevalence of drug-related hospital admissions, ranging from 1% to 41% in a recent systematic review. Interestingly, only three out of 16 studies included in that review reported kappa values for independent assessors and gave their professional background. Further information about the reviewers, such as number of years of experience, was not provided.

Strengths of the present study include its contribution to the sparse evidence base regarding the reliability of clinically relevant drug treatment assessments. Indeed, prior studies reporting on the inter-rater agreement on quality of drug treatment have concerned PIMs/PPOs. These are surrogate variables with limited ability to reflect drug treatment quality. Further, a few studies have reported on reliability between at least three assessors, with either pharmaceutical

### Table 4

| Reason for admission | Drug treatment suspected of causing admission (main/contributory cause) | Relationship between drug treatment and admission | Admission related to prescribing error |
|----------------------|------------------------------------------------------------------------|-------------------------------------------------|--------------------------------------|
|                      |                                                                        | Individual assessments                           | Individual assessments               |
|                      |                                                                        | Main reason (n)                                  | Contributory reason (n)              |
|                      |                                                                        | Not related (n)                                  | Consensus assessment                |
|                      |                                                                        | Definitely (n)                                   | Possibly (n)                        |
|                      |                                                                        | No (n)                                           | Consensus assessment                |
| Neutropenic fever    | Chemotherapy                                                          | 5                                                | 0                                   |
|                      |                                                                        | 0                                                | M                                   |
|                      |                                                                        | 0                                                | 1                                   |
|                      |                                                                        | 4                                                | N                                   |
| Abdominal pain       | Inadequate laxatives                                                  | 1                                                | 3                                   |
| with constipation     |                                                                        | 1                                                | C                                   |
|                      |                                                                        | 1                                                | P                                   |
| Subdural haematoma   | Rivaroxaban                                                           | 0                                                | 4                                   |
|                      |                                                                        | 1                                                | C                                   |
|                      |                                                                        | 0                                                | 2                                   |
|                      |                                                                        | 3                                                | N                                   |
| Hypotension, nausea, | Metoprolol, irbesartan, oxycodone                                     | 0                                                | 5                                   |
| pain                 |                                                                        | 0                                                | 0                                   |
|                      |                                                                        | 5                                                | N                                   |
| Fall, head injury    | Hydroxyzine, diazepam, alprazolam, zolpidem                           | 0                                                | 4                                   |
|                      |                                                                        | 1                                                | C                                   |
|                      |                                                                        | 1                                                | 3                                   |
|                      |                                                                        | 1                                                | P                                   |
| Erysipelas, reduced  | Metolazone, furosemide                                                | 0                                                | 3                                   |
| renal function       |                                                                        | 2                                                | C                                   |
|                      |                                                                        | 0                                                | 1                                   |
|                      |                                                                        | 4                                                | P                                   |
| Pyelonephritis, renal | Candesartan                                                            | 0                                                | 4                                   |
| failure              |                                                                        | 1                                                | C                                   |
|                      |                                                                        | 0                                                | 3                                   |
|                      |                                                                        | 2                                                | N                                   |
| RSV infection        | Methotrexate                                                           | 0                                                | 1                                   |
|                      |                                                                        | 4                                                | C                                   |
|                      |                                                                        | 0                                                | 1                                   |
|                      |                                                                        | 4                                                | N                                   |
| Renal failure, worsened | Furosemide, no amiodaron (recently discontinued)                 | 0                                                | 3                                   |
| heart failure        |                                                                        | 2                                                | C                                   |
|                      |                                                                        | 0                                                | 2                                   |
|                      |                                                                        | 3                                                | N                                   |

C, contributory reason; D, definitely a prescribing error; M, main reason; N, not a prescribing error; P, possibly a prescribing error; RSV, respiratory syncytial virus.
or medical expertise, regarding ADRs, but none of these used a random sample of hospital patients.\textsuperscript{22–25,27,43} The assessed cases were either pre-selected based on being “drug-related”,\textsuperscript{22–25} or they were elective patients\textsuperscript{43} or fictitious.\textsuperscript{27} Another strength of the present study is that the assessments were systematically preceded by the application of extensive screening and assessment tools, so as not to omit important drug-related aspects. Nevertheless, this approach may have contributed to the strong to almost perfect agreement between the assessors regarding drug treatment quality issues, and further research is thus warranted to elucidate the inter-rater agreement without such standardisation.

An important limitation of this study is the small number of patients assessed. Hence, conclusions regarding potential differences between groups of assessors cannot be drawn. Furthermore, we cannot exclude that the results, in particular in the resident group, were affected by experience of individual assessors. In future research, evidence regarding potential differences between assessors with different level of expertise could be achieved by increasing the number of assessors and patients. It may also be regarded as a limitation that the assessors had no previous experience of drug treatment assessment in the research setting. However, all assessors were used to dealing with patients, including their drug treatment, in their daily work as internists and future internists. Another limitation is that the underlying information for the assessments was restricted to health records. Indeed, all relevant information about the medical history, diagnoses, examinations, and laboratory tests may not be fully documented. Yet this data source is more comprehensive than drug registers, sometimes linked to other registers, which are frequently used in studies on prescribing practices.\textsuperscript{44}

In summary, this study illustrates that the reliability of assessments of prescribing quality, as well as adverse consequences, is a matter of methodological concern: although the inter-rater agreement between medically experienced physicians is in general above the desirable level of 0.80 for quality-related aspects, weak to moderate agreement can be expected for ADR-related issues. For physicians at an earlier stage of their medical career, assessments of ADR-related issues seem to diverge to a greater extent than assessments of quality-related aspects. Previous research results on prescribing practices and drug-related admissions need to be interpreted with this in mind. To yield acceptably reliable results, our results suggest that having two residents to capture quality of drug treatment may be sufficient, while specialist physicians could be involved in the assessments of ADRs.

ETHICAL APPROVAL
The study was approved by the Regional Ethical Review Board in Gothenburg (ID: 463-18).

ACKNOWLEDGEMENTS
This work was supported by grants from the Swedish Research Council (2013-02639) and the Swedish state under the ALF agreement between the Swedish government and the county councils (ALFGBG-716941). The funding source did not influence the design or conduct of the study, nor the collection, management, analysis or interpretation of the data, or the preparation, review and approval of the manuscript.

COMPETING INTERESTS
The authors declare that they have no conflict of interest.

CONTRIBUTORS
J.L. and S.M.W. conceived and designed the study. J.L. prepared background data and, together with L.H., N.D.A., E.P. and P.T., performed the individual quality assessments and consensus discussion. J.L. and S.M.W. drafted the manuscript and all authors revised it for intellectual content. J.L. is the guarantor of this work.

DATA AVAILABILITY STATEMENT
The data sets generated and analysed during the current study are not publicly available because of Swedish data protection laws. The data can only be shared with authorised persons after approval from the Swedish Ethical Review Authority (https://etikprovningsmyndigheten.se).

ORCID
Johan Lönnbro \textsuperscript{https://orcid.org/0000-0001-6917-1076}
Susanna M. Wallerstedt \textsuperscript{https://orcid.org/0000-0001-7238-1680}

REFERENCES
1. Schmidt SJ, Wurmback VS, Lampert A, et al. Individual factors increasing complexity of drug treatment—a narrative review. Eur J Clin Pharmacol. 2020;76(6):745-754.
2. Guaraldo L, Cano FG, Damasceno GS, Rozenfeld S. Inappropriate medication use among the elderly: a systematic review of administrative databases. BMC Geriatr. 2011;11(1):79.
3. Tommelein E, Mheuys E, Petrovic M, Somers A, Colin P, Boussery K. Potentially inappropriate prescribing in community-dwelling older people across Europe: a systematic literature review. Eur J Clin Pharmacol. 2015;71(12):1415-1427.
4. Rankin A, Cadogan CA, Patterson SM, et al. Interventions to improve the appropriate use of polypharmacy for older people. Cochrane Database Syst Rev. 2018;9:CD008165.
5. Allred DP, Raynor DK, Hughes C, Barber N, Chen TF, Spoor P. Interventions to optimise prescribing for older people in care homes. Cochrane Database Syst Rev. 2013;2:CD009095.
6. Lewis PJ, Dorman T, Taylor D, Tully MP, Wass V, Ashcroft DM. Prevalence, incidence and nature of prescribing errors in hospital inpatients: a systematic review. Drug Saf. 2009;32(5):379-389.
7. Pirmohamed M, James S, Meakin S, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. BMJ. 2004;329(7456):15-19.
8. Pedros C, Formiga F, Corbella X, Arnau JM. Adverse drug reactions leading to urgent hospital admission in an elderly population: prevalence and main features. Eur J Clin Pharmacol. 2016;72(2):219-226.
9. Nwya K, Sri Sai Kiran V, Ragoo N, Jayaprakash B, Sonal Sekhar M. Systemic review on drug related hospital admissions—A PubMed based search. Saudi Pharm J. 2015;23(1):1-8.
10. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA. 1998;279(15):1200-1205.
11. Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? BMJ. 2014;348:g3725.
12. Wallerstedt SM, Belfrage B, Fastbom J. Association between drug-specific indicators of prescribing quality and quality of drug treatment: a validation study. *Pharmacoepidemiol Drug Saf*. 2015;24(9):906-914.
13. Belfrage B, Koldestam A, Sjöberg C, Wallerstedt SM. Number of drugs in the medication list as an indicator of prescribing quality: a validation study of polypharmacy indicators in older hip fracture patients. *Eur J Clin Pharmacol*. 2015;71(3):363-368.
14. Lönnbro J, Wallerstedt SM. Clinical relevance of the STOPP/START criteria in hip fracture patients. *Eur J Clin Pharmacol*. 2017;73(4):499-505.
15. Wallerstedt S, Belfrage B, Koldestam A, Sjöberg C. Many patients with suboptimal drug treatment are missed. Sensitivity, specificity and prediction value of drug-specific quality indicators for the elderly. *Läkartidningen*. 2016;113:DSU4.
16. Johansson ML, Hägg S, Wallerstedt SM. Impact of information letters on the reporting rate of adverse drug reactions and the quality of the reports: a randomized controlled study. *BMC Clin Pharmacol*. 2011;11(1):14.
17. World Health Organization (WHO)–Uppsala Monitoring Centre (UMC). The WHO-UMC system for standardized case causality assessment. https://www.who-umc.org/media/164200/who-umc-causality-assessment-new-logo.pdf. Accessed January 27, 2021.
18. Lavan AH, Gallagher P, O’Mahony D. Inter-rater reliability of STOPP/FRail [Screening Tool of Older Persons Prescriptions in Frail adults with limited life expectancy] criteria amongst 12 physicians. *Eur J Clin Pharmacol*. 2018;74(3):331-338.
19. Marcum ZA, Amuon ME, Hanlon JT, et al. Prevalence of unplanned hospitalizations caused by adverse drug reactions in older veterans. *J Am Geriatr Soc*. 2012;60(1):34-41.
20. Woo SA, Cragg A, Wickham ME, et al. Methods for evaluating adverse drug event preventability in emergency department patients. *BMC Med Res Methodol*. 2018;18(1):160.
21. Hakkarainen KM, Andersson Sundell K, Petzold M, Hägg S. Methods for assessing the preventability of adverse drug events: a systematic review. *Drug Saf*. 2012;35(2):105-126.
22. Howard RL, Avery AJ, Howard PD, Partridge M. Investigation into the reasons for preventable drug related admissions to a medical admissions unit: observational study. *Qual Saf Health Care*. 2003;12(4):280-285.
23. Olivier P, Boulbès O, Tubery M, Laque D, Montastruc JL, Lapeyre-Mestre M. Assessing the feasibility of using an adverse drug reaction preventability scale in clinical practice: a study in a French emergency department. *Drug Saf*. 2002;25(14):1035-1044.
24. Ahern F, Sahn LJ, Lynch D, McCarthy S. Determining the frequency and preventability of adverse drug reaction-related admissions to an Irish University Hospital: a cross-sectional study, *Emerg Med J*. 2014;31(1):24-29.
25. Tangiisuran B, Auyeung V, Cheek L, Rajkumar C, Davies G. Inter-rater reliability of the assessment of adverse drug reactions in the hospitalised elderly, *J Nutr Health Aging*. 2013;17(8):700-705.
26. O’Mahony D, O’Sullivan D, Byrne S, O’Connor MN, Ryan C, Gallagher P. STOPP/START criteria for potentially inappropriate prescribing in older people: version 2. *Age Ageing*. 2015;44(2):213-218.
27. Thevelin S, Spinewine A, Beuscart JB, et al. Development of a standardized chart review method to identify drug-related hospital admissions in older people. *Br J Clin Pharmacol*. 2018;84(11):2600-2614.
28. Hallas J, Harvald B, Gram LF, et al. Drug related hospital admissions: the role of definitions and intensity of data collection, and the possibility of prevention. *J Intern Med*. 1990;228(2):83-90.
29. Fastbom J, Johnell K. National indicators for quality of drug therapy in older persons: the Swedish experience from the first 10 years. *Drugs Aging*. 2015;32(3):189-199.
30. Dean B, Barber N, Schachter M. What is a prescribing error? *Qual Health Care*. 2000;9(4):232-237.
31. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol*. 2008;61(1):29-48.
32. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med* (Zagreb). 2012;22(3):276-282.
33. Rotondi MA, Donner A. A confidence interval approach to sample size estimation for interobserver agreement studies with multiple raters and outcomes. *J Clin Epidemiol*. 2012;65(7):778-784.
34. Varmdal T, Ellekjær H, Fjærtoft H, Indredavik B, Lydersen S, Bonaa KH. Inter-rater reliability of a national acute stroke register. *BMC Res Notes*. 2015;8(1):584.
35. Björk J, Grubb A, Sterner G, Nyman U. Revised equations for estimating glomerular filtration rate based on the Lund-Malmö Study cohort. *Scand J Clin Lab Invest*. 2011;71(3):232-239.
36. Schurig AM, Böhme M, Just KS, et al. Adverse drug reactions (ADR) and emergencies. *Dtsh Arztebl Int*. 2018;115(15):251-258.
37. Brilleman SL, Salisbury C. Comparing measures of multimorbidity to predict outcomes in primary care: a cross sectional study. *Fam Pract*. 2013;30(2):172-178.
38. Repp KL, Hayes C 3rd, Woods TM, Allen KB, Kennedy K, Borkan MA. Drug-related problems and hospital admissions in cardiac transplant recipients. *Ann Pharmacother*. 2012;46(10):1299-1307.
39. van Doornmaal JE, Mol PG, van den Bemt PM, et al. Reliability of the assessment of preventable adverse drug events in daily clinical practice. *Pharmacoepidemiol Drug Saf*. 2008;17(7):645-654.
40. Beuscart J-B, Knol W, Cullinan S, et al. International core outcome set for clinical trials of medication review in multi-morbid older patients with polypharmacy. *BMC Med*. 2018;16(1):21.
41. Rankin A, Cadogan CA, Ryan C, Clyne B, Smith SM, Hughes CM. Core outcome set for trials aimed at improving the appropriateness of polypharmacy in older people in primary care. *J Am Geriatr Soc*. 2018;66(6):1206-1212.
42. Ayawef MB, Tegegn HG, Abdela OA. Drug related hospital admissions: a systematic review of the recent literatures. *Bull Emerg Trauma*. 2019;7(4):339-346.
43. de Boer M, Kiewiet JJS, Boeker EB, et al. A targeted method for standardized assessment of adverse drug events in surgical patients. *J Eval Clin Pract*. 2013;19(6):1073-1082.
44. Wallerstedt SM, Wettermark B, Hoffmann M. The first decade with the Swedish prescribed drug register—a systematic review of the output in the scientific literature. *Basic Clin Pharmacol Toxicol*. 2016;119(5):464-469.

**How to cite this article:** Lönnbro J, Holmqvist L, Persson E, Thysell P, Åberg ND, Wallerstedt SM. Inter-rater reliability of assessments regarding the quality of drug treatment, and drug-related hospital admissions. *Brit Jnl Clinical Pharma*. 2021;1–10. [https://doi.org/10.1111/bcp.14790](https://doi.org/10.1111/bcp.14790)