



# Strategies and Challenges in Coding Ambiguous Information Using MedDRA®: An Exploration Among Norwegian Pharmacovigilance Officers

Tahmineh Garmann<sup>1</sup> · Hilde Samdal<sup>2</sup> · Daniele Sartori<sup>3</sup> · David Jahanlu<sup>1</sup> · Fredrik Andersen<sup>4</sup> · Elena Rocca<sup>1</sup>

Accepted: 4 June 2025 / Published online: 1 July 2025

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025

## Abstract

**Introduction** The Medical Dictionary for Regulatory Activities (MedDRA®) is an international standardized medical terminology used to code various types of medical information, including safety reports of suspected adverse reactions to medicines. Quantitative studies have highlighted varying levels of coding inconsistency across MedDRA®-relevant platforms, though the possible grounds of such inconsistency remain unclear.

**Objective** We explored the reasoning and strategies employed by pharmacovigilance officers when coding selected ambiguous adverse events to MedDRA®, categorized the types of coding inconsistencies, and explored sources of the inconsistencies.

**Methods** Pharmacovigilance officers from the Norwegian public health sector were invited to participate in a survey-based, cross-sectional study followed by focus group interviews. The survey consisted of 11 coding tasks, with varying degrees of ambiguity, purposively sampled from the Norwegian pharmacovigilance registry. Participants selected the appropriate MedDRA® terms and graded the difficulty level of each task on a scale from 1 (least difficult) to 4 (most difficult). Terms selected by participants were compared with a Standard Term Selection (STS), agreed upon by the authors in consultation with a MedDRA® trainer. Inconsistencies with the STS were classified as omission (missing term), substitution (extra term selected in the presence of an omission), and addition (extra term selected and none omitted). In focus groups, participants discussed challenges in the coding tasks and the strategies they used to overcome them. Interview transcripts were analyzed using thematic analysis.

**Results** In total, 26 coders (79% of the eligible population) completed the survey. Of the survey answers, 36% were identical to the STS; answers consistent with the STS varied across the specific coding tasks and did not align with the perceived difficulty of the tasks. The most common inconsistency (30% of the survey answers) arose from substituting one of multiple MedDRA® terms. Of the survey answers, 18% included omissions without substitutions, and 6% added unnecessary terms to the STS. Eight of the 26 coders (31%) participated in the focus group interviews. Focus group themes revealed that substitutions were explained by difficulties in translating lay language to medical terminology, finding accurate English translations for Norwegian medical terms, and fitting complex descriptions into MedDRA® terms. This was explained by themes related to ambiguity-resolution strategies. Themes explaining omissions included strategies for resolving ambiguity, contextual thinking, causal and pharmacological reasoning in the coding process, and information categorization.

**Conclusions** Tailored training programs and clear institutional guidelines are needed to target the sources of coding inconsistencies suggested by this study.

## 1 Introduction

Pharmacovigilance, the science of detecting, assessing, and preventing possible adverse reactions to medicines and any other drug-related problem, primarily relies on what is known as post-marketing “passive safety monitoring.” This

process typically involves the reporting of possible adverse reactions to medicines, through standardized reporting forms, to a pharmacovigilance database. In Norway, reports are processed and coded by pharmacovigilance officers at the Norwegian Medical Products Agency (NOMA), the Regional Medicines Information and Pharmacovigilance Centers (RELIS), and the Norwegian Institute of Public Health (NIPH) [1–3]. The recording of safety reports in pharmacovigilance databases involves digitally transcribing

Extended author information available on the last page of the article

### Key Points

This study explored the reasoning and strategies used by Norwegian pharmacovigilance officers to code ambiguous medical information using MedDRA<sup>®</sup> and described common inconsistencies when coding selected tasks.

The results showed that 36% of the coders' answers matched a standard reference with the most common mistake being the substitution of terms. Additionally 18% of the answers omitted terms without substitutions and 6% added unnecessary terms. The coding consistency did not match the perceived difficulty of each task.

Key sources of ambiguity emerging from focus group discussions included translating lay language to medical terms finding accurate English translations contextual thinking in the coding process and variability in handling ambiguous information.

The findings suggest that specific training programs and clear institutional guidelines are essential to help coders overcome these challenges and improve the accuracy of MedDRA<sup>®</sup> coding.

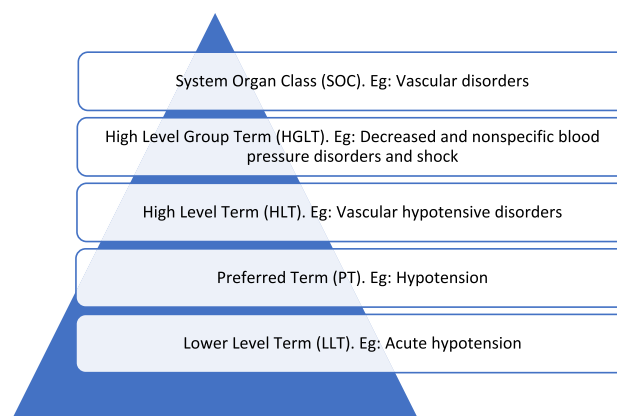
and coding the reported information using standardized international terms for medicines and events [1–3].

The Medical Dictionary for Regulatory Activities (MedDRA<sup>®</sup>) is an international standardized medical terminology developed in the 1990s under the auspices of the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use and managed by the MedDRA<sup>®</sup> Maintenance and Support Services Organization [4]. It is a hierarchical system consisting of five levels, from general groupings to higher specificity: system organ class (SOC), high-level group term (HGLT), high-level term (HLT), preferred term (PT), and lowest-level term (LLT) (Fig. 1).

Adverse events are coded to LLTs and retrieved for data analyses at higher levels of the hierarchy. To facilitate data retrieval, PTs can be assigned to multiple SOC, one being primary and—where appropriate—one or more secondary. For instance, the PT “influenza” is included both in the primary SOC “infections and infestations” and in the secondary SOC “respiratory, thoracic, and mediastinal disorders.” Data retrieval for conditions of interest is also possible through standardized MedDRA<sup>®</sup> queries, which are group sets of clinically related terms [5]. The granularity of MedDRA<sup>®</sup> (more than 78,000 current LLTs in version 26.1) affords flexibility for retrieval and analysis but also the potential for inconsistent coding [3, 6–8]. As a result, the way in which

MedDRA<sup>®</sup> coding guidelines are applied by different coders may vary. Research indicates that automated systems using natural language processing can be useful in assisting the process of coding suspected adverse drug reactions, but human review is still essential [9]. Ultimately, the way reported information is coded to MedDRA<sup>®</sup> affects data retrieval from the database, statistical analyses, and clinical evaluation of potential signals [3].

Best practice guidelines for coding are provided in the MedDRA<sup>®</sup> Term Selection: Points to Consider (MTS:PTC) document, which is updated annually [8]. These guidelines offer recommendations for term selection, in some cases with a preferred and alternate approach. Thus, term selection may vary between organizations, although an organization should always be consistent in their chosen approach to coding similar reported events and are encouraged to keep a record of their internal coding conventions. Tonéatti et al. [10] investigated coding discrepancies between two trained MedDRA<sup>®</sup> coders who were tasked with coding events reported in HIV clinical trials. Coders were also asked to assess the difficulty of the verbatims or events they were coding. The findings were subsequently reviewed by a coding committee, revealing a 12% rate of coding inaccuracies. A greater proportion of these inaccuracies occurred in the verbatims that the coders had rated as difficult, suggesting that the perceived difficulty of adverse event descriptions influences the accuracy of the coding process [10]. In 2018, US Food and Drug Administration experts compared over 3000 MedDRA<sup>®</sup> codes assigned by patient representative groups to patient-reported text against those assigned by regulatory authorities. Less than 3% of the MedDRA<sup>®</sup> codes chosen by patients diverged from the regulators' choices. When codes diverged, patients assigned more general codes, whereas regulators used specific codes [11]. In a study concerning reports of adverse events with COVID-19 vaccines,



**Fig. 1** Illustrative example of MedDRA<sup>®</sup> hierarchy. *HGLT* high-level group term, *HLT* high-level term, *LLT* lower-level term, *PT* preferred term, *SOC* system organ class

half of the 1500 reported terms (verbatimims) coded to an LLT subsumed by the HLT “medication errors and other product use errors and issues” were unspecific or split-coded [12].

All in all, evidence indicates that there are inconsistencies in the application of MedDRA<sup>®</sup> guidelines and that it may be difficult to code events that are poorly described or ambiguous. To our knowledge, there is no qualitative analysis of the strategies coders use to select codes for capturing ambiguous information. Such qualitative knowledge may help design effective training programs or improve existing guidelines.

The primary aim of this study was to explore how Norwegian pharmacovigilance officers used MedDRA<sup>®</sup> to code information about adverse events in ambiguous verbatimims, with the following objectives:

- (1) Describe the level of consistency among coders and between coders and a MedDRA<sup>®</sup> standard PT selection.
- (2) Describe the types of coding inconsistencies, if any.
- (3) Explore the strategies and resources that coders used when applying MedDRA<sup>®</sup> guidelines to ambiguous verbatimims.

A secondary aim of this study was to outline a general procedure for assessing MedDRA<sup>®</sup> coding practices within a specific user group, for the purpose of capacity building.

## 2 Methods

### 2.1 Selection of Study Participants and Recruitment

Eligible participants included all the MedDRA<sup>®</sup> coders working in the public sector in Norway (also known as pharmacovigilance officers). At the time of the study, the eligible study population consisted of 33 coders, employed by NOMA ( $n = 8$ ), RELIS ( $n = 20$ ), and NIPH ( $n = 5$ ). The leaders of the three institutions were approached via email and sent a link to *Nettskjema*, an online survey platform operated by the University Information Technology Center at University of Oslo. Through the online link, they were sent recruitment materials and asked to forward them to MedDRA<sup>®</sup> coders. Coders who were interested in participating in the focus groups upon completion of the anonymous survey were invited to send their contact information to the project leader. Coders were informed that participation was voluntary.

### 2.2 Survey Design

The survey presented 11 “verbatimims” (event terms as reported) extracted from individual case safety reports in the

Norwegian Adverse Drug Reaction Registry (NorADRR) (hereafter referred to as coding tasks). Participants were asked to code them to LLT as open-ended answers using the version of MedDRA<sup>®</sup> in use in Norway: English MedDRA<sup>®</sup> version 26.1 (Table 1). The open-ended format allowed for unrestricted coding choices, without suggestions or pre-selected codes from which to choose. Upon completion of each coding task, participants were asked to grade the level of difficulty on a scale from 1 (easy) to 4 (very difficult). In line with the study aim, the survey was designed so that coding tasks would fulfil two requirements: (1) ambiguity: the coding tasks had to contain some elements that are known to be challenging for MedDRA<sup>®</sup> coders; (2) plausibility: tasks had to be realistic and should be recognized by participants as plausible job tasks. Purposive sampling of all NorADRR verbatimims between 2018 and 2020 was carried out to ensure the requirements were fulfilled. Briefly, the authors compiled a list of frequent sources of coding variation, based on MedDRA<sup>®</sup> coding training experience (see Table 1 in the electronic supplementary material [ESM]). The authors subsequently skimmed the NorADRR verbatimims using the list’s items as search keys, screening for cases that exemplified MTS:PTC but remained ambiguous enough to provide a challenge for the coders. In addition to the coding tasks, the survey included multiple choice questions about the participants’ educational background, type of MedDRA<sup>®</sup> training, years of work experience with MedDRA<sup>®</sup>, and MedDRA<sup>®</sup> coding routines. No directly identifying information (such as name, address, email contact, or date of birth) were collected.

### 2.3 Focus Group Interviews Design and Execution

Participants who accepted the invitation to participate in 60-min, audio-recorded focus group interviews were assigned to groups based on time availability. In total, three focus group interviews were led by TG and ER about 1 month after survey completion. Recordings were transcribed at the end of each focus group. The interview guide (see Table 2 in the ESM) was informed by a preliminary analysis of survey results. Coding tasks that yielded low consistency and/or high perceived degree of difficulty were re-proposed to the group and used to generate discussion about strategies adopted when facing ambiguity. In addition, perspectives about training and challenges with MedDRA<sup>®</sup> coding were asked independently from coding tasks. Lastly, the survey’s concept validity was tested by asking about the extent to which the survey tasks were perceived as ambiguous and plausible. Answers to questions on concept validity were considered to adjust the survey analysis (see Sect. 2.4).

**Table 1** Complete list of the coding tasks used in the study

N	Survey coding tasks	Standard PT selection
1 (S)	Itching on palms, soles of feet and on the chest. No rash	Pruritus
2 (S)	Feeling down and moody, unusually tired even though I get to sleep enough, worn out, and little energy after starting with migraine medication	Depressed mood, somnolence, fatigue, asthenia, mood swings
3 (S, FG)	Worsening of low metabolism with increase in anti TPO	Hypometabolism, anti-thyroid antibody increased, condition aggravated
4 (S)	Serious hypotension after introducing the anesthesia due to laparoscopic cholecystectomy	Procedural hypotension
5 (S)	Young girl receiving hormonal intrauterine contraceptive develops lower abdominal cramps. 5 weeks after the insertion during a control, they found out that the device is not in the right location. The patient developed perforation of the uterus, the device was loose in the abdomen and had to be removed through laparoscopy	Abdominal pain lower, device dislocation, uterine perforation
6 (S, FG)	Intracerebral hemorrhage after patient admitted for fall/syncope. Not sure if there was head trauma	Syncope, fall, cerebral hemorrhage, head injury
7 (FG)	Daily headache and dizziness. Many purple lumps on the chest, and a spot on the forehead that resembles skin cancer	Agreement not reached by authors
8 (S)	Pregnant woman used antiepileptic medicine (valproate 600 + 450 mg) throughout her pregnancy. The patient has motoric, social, and linguistic developmental delay	Exposure during pregnancy, neurodevelopmental delay, motor developmental delay, speech disorder developmental
9 (S, FG)	Mother is pregnant 6 months after the last dose of acne medicine (isotretinoin: A-vitamin derivative). During pregnancy screening on week 19+3, a serious heart failure is observed, and the pregnancy was terminated. Autopsy results of the fetus confirmed the screening results	Fetal cardiac disorder, maternal exposure before pregnancy, abortion induced
10 (S, FG)	Water in the body, especially swelling on top of the cheek and around the eyes. I checked my hands; they were also swollen	Periorbital swelling, peripheral swelling, swelling face
11 (S, FG)	The patient used a new supplement for about 2 months, developed elevated liver enzymes (all of them), including INR increased (cell markers ALT/ASAT). Thrombocyte count decreased. Symptoms improved after medication was discontinued. Patient got hospitalized and it turns out that she/he has been using alcohol along with the medications	Internal normalized ratio increased, aspartate aminotransferase increased, alanine aminotransferase increased, platelet count decreased
12 (S)	Patient diagnosed with ca renis cum met in July 2020; Clear cell kidney carcinoma	Clear cell renal cell carcinoma (LLT: Clear cell renal cell carcinoma metastatic)

Column 1 indicates whether the tasks were assigned in form of survey (S), focus group interviews (FG) or both. Column 2 shows the complete list of the coding tasks used in the study. Column 3 indicates, in correspondence of each coding task, the standard PT selection as agreed by the authors in consultancy with the MSSO. The shaded cells indicate the baseline task

ALT alanine transaminase, AST aspartate aminotransferase, INR international normalized ratio, LLT lower-level term, MSSO MedDRA® Maintenance and Support Services Organization, PT preferred term, TPO thrombopoietin, TSP term selection principles

<sup>a</sup>MedDRA TSPs of reference were TSP 3.7.3. "Event reported at more body sites should be reported jointly if they link to the same PT" (task 1); TSP 2.10. "Select terms for all reported information, do not add information" (task 2); TSP 3.9 Modification of pre-existing conditions: "Select a term for the pre-existing condition and second term for the modification for the condition" (task 3); TSP 2.4. "Select the LLT that most accurately reflects the reported verbatim" (tasks 4, 11, 12); TSP 3.19.1. "If available, select a term that reflects both the device-related event and the clinical consequence" (task 5); TSP 3.1. "The preferred option for a single or multiple provisional diagnosis(es) is to select a term(s) for the diagnosis(es) and terms for reported signs and symptoms. This is because a provisional diagnosis may change while signs/symptoms do not" (tasks 6, 7); TSP 3.10.2, Exposures during pregnancy and breastfeeding, Events in the child or fetus: "Select terms for both the type of exposure and any adverse event" (tasks 8, 9); TSP 3.5.4. "If splitting the reported AE provides more clinical information, select more than one term (task 10); TSP 3.7.3. "Event reported at more body sites should be reported jointly if they link to the same PT" (task 10); TSP 3.20.2. "Two products may be used together, but if the reporter does not specifically state that an interaction has occurred, select terms only for the medical event reported" (task 11)

## 2.4 Analysis of Survey Results

The authors reviewed the coding tasks and discussed which LLTs should be selected to comply with the MTS:PTC guideline for standard LLT selection. Tasks were first coded individually, and inconsistencies were addressed together and in consultation with a MedDRA<sup>®</sup> trainer. For one of the 12 coding tasks initially selected (coding task #7 in Table 1), no standard LLT selection could be found because of the high level of ambiguity. It was therefore decided to exclude task 7 from the analysis and include it in the focus group interview guide. The LLTs assigned by the participants and the standard LLTs assigned by the author group were converted to the corresponding PTs (standard PT selection, see Table 1), and the analyses were carried out at PT level. The standard PT selection consisted of 32 unique PTs.

Terms that could be classified other than “adverse events” were excluded from the analyses. Excluded PTs related to:

- medication history/suspected medicines: migraine prophylaxis, induction of anesthesia, anesthesia procedure, anesthesia (task 2); intrauterine contraception (task 5), intrauterine contraceptive device insertion (task 5); pre-medication (task 11).
- medical history: cholecystectomy (task 4), laparoscopy (task 4); alcohol use (task 11).
- investigations: autopsy (task 9).
- action taken: laparoscopic surgery, intrauterine contraceptive device removal (task 5); therapy cessation (task 11).
- indication: migraine (task 2).
- seriousness: hospitalization (task 6).

This exclusion was informed by a concept validity issue detected during the focus groups. The issue originated because coders in normal practice assign LLTs to structured fields, which provide categories to what they code, such as an adverse event, medical history, or indication. However, the experimental settings (i.e. the survey) admitted only free-text answers, thus lacking the context they were habitually afforded. Focus group participants noted that the coding tasks could be interpreted in two ways: (1) indicate the LLTs that would, in normal practice, belong to the structured field for “adverse event” or (2) indicate the LLTs for all the information in the verbatims, regardless of whether, in normal practice, they would belong to structured fields other than “adverse event.” This two-fold interpretation of the coding tasks could have given rise to some of the variations in the answers. To avoid this, we excluded from the conclusive survey analysis the variation in LLTs that, according to focus group informants, could fit in structured fields other than “adverse events”.

The level of consistency among survey participants was defined as the frequency of the PT combination that was selected by most participants (operationalized as the maximum number of identical answers per coding task). The level of consistency with the standard PT selection was operationalized in two ways: (1) number of answers identical to the standard PT selection and (2) number of answers containing the complete standard PT selection but formally deviating from it because of additional PTs.

The type of inconsistency between survey participants and the standard PT selection was categorized as follows:

- omission: one of the PTs in the standard PT selection is missing
- substitution: in the presence of omission(s), an extra PT with respect to the standard PT selection is selected
- addition: an extra PT is selected with respect to the complete standard PT selection.

For examples, see Table 2.

Repeated inconsistencies by the same coder on the same set of verbatims could be counted more than once. For instance, in the example in Table 2, an answer consisting of “periorbital swelling” would be counted as two omissions. To better describe the variation in coding, we categorized and quantified the answers that lacked one or more standard PT(s) into answers containing no substituting PTs, answers containing both standard PTs and substituting PTs, and answers containing only substituting PTs.

To evaluate the potential impact of each substitution in subsequent data retrieval from the registry, hierarchy analyses were performed to see whether the substitution was reflected at a higher level of the MedDRA<sup>®</sup> hierarchy. If the assigned terms were multiaxial, all branches were included in the analysis. Variations were categorized into three levels: (1) substitution at PT level, (2) substitution at HLT level, and (3) substitution at SOC level.

## 2.5 Analysis of Focus Group Data

TG and ER used reflexive thematic analysis, independently splitting transcripts into minor content units (codes) and subsequently grouped codes together into general themes [13]. This strategy helps organize statements and discover patterns between statements that appear in different interviews and/or at different moments within interviews. Then, TG, FA, and ER compared, discussed, and enriched their respective codes when they differed. Codes were then grouped into potential themes, and potential themes were evaluated together to see whether the data cohered meaningfully. Finally, TG, FA, and ER discussed and revised the candidate themes.

**Table 2** Example illustrating the categorization of possible inconsistencies between survey answers and the standard preferred term (PT) selection

Standard PT selection	Survey answer	Category of inconsistency
Periorbital swelling, swelling face, peripheral swelling	Periorbital swelling, swelling face	Omission
	Periorbital swelling, swelling face, peripheral swelling, edema	Addition
	Periorbital swelling, swelling face, generalized edema	Omission, substitution

## 2.6 Research Ethics

This research was carried out in accordance with relevant guidelines and regulations. The regional research ethics committee south-east Norway concluded that the project fell outside the scope of the Norwegian Health Research Act (reference number 539798, 17.04.2023). Data collection and storage were designed in consultation with the Norwegian Agency for Shared Services in Education and Research to ensure adherence to data protection and privacy regulations (reference number 766729, 10.05.2023).

## 3 Results

### 3.1 Survey Results

In total, 26 codes completed the survey (79% of the total study population), one of whom did not indicate their educational background. The participation rate per institution was 100% (5/5) for NIPH, 75% (6/8) for NOMA, and 75% (15/20) for RELIS. An overview of the participant demographics is presented in Table 3.

The level of coding consistency among participants, the level of consistency with the standard PT selection, and the number of answers with low perceived difficulty per task are shown in Fig. 2. In the figure, the sequence of the tasks is reorganized in ascending order of low perceived difficulty. The data in Fig. 2 show that, as expected, all the answers to the baseline task were identical to the standard. For the other tasks, the answers containing the complete standard PT selection (with or without addition) ranged from a maximum of 25 (task 12) to a minimum of 2 (task 6). For 8 of 11 tasks, the largest agreement corresponded to the standard PT selection. Interestingly, task 3 had the third largest maximum agreement (15 answers) but no answer containing the standard PT selection. The lowest agreement was recorded in task 10 (three answers), which, by contrast, had the lowest perceived difficulty. Also, task 2 had the same perceived difficulty as the baseline but scored low, both for answers containing the standard PT selection (three answers) and for the size of the largest agreement (five answers). Hence,

the data suggest that the perceived difficulty did match the coding performance in our selection of tasks.

Table 4 shows the categorization of coding omissions into answers containing no substituting PTs, answers containing both standard PTs and substituting PTs, and answers containing only substituting PTs. The latter category represented the only possible deviation for tasks with one PT in the standard PT selection. Among these, the standard PT in task 4 (procedural hypotension) was substituted by 12 participants, whereas the standard PT in task 12 (clear cell renal cell carcinoma) was substituted by only one participant. However, for task 12, it is worth noting that, although the selected PT was consistent in 25 cases, only 16 of the participants captured the additional information “metastatic carcinoma,” something that is possible to observe only at the LLT level and therefore not visible in our quantification.

It was generally more common for coders to substitute a part of the standard PTs in tasks in which the standard PT selection contained three, four, or five PTs than it was for tasks with only one PT in the standard PT selection. However, the data show no clear increasing or decreasing trend. For instance, in task 2, which contains five PTs in the standard PT selection, coders commonly omitted one PT among “somnia, fatigue, asthenia” without substituting it. Further, in task 6, which contained four PTs in the standard selections, only one of 24 participants who did not code the information regarding fall and head injury chose a substituting PT. In our selection of tasks, heterogeneity did not seem to be affected by the number of verbatims, rather by the type of information in the verbatim.

Of a total 286 survey tasks answered by the 26 participants, 103 (36%) were identical to the standard PT selection. Figure 3 shows the distribution of the remaining 183 survey answers, which were inconsistent with the standard PT selection, into the different types of identified inconsistencies, as well as the cumulative impact of each type of inconsistency. The most frequent deviation from the standard PT selection was omission(s) and use of one substituting PT (88 answers), followed by omission(s) without any substitution (53 answers). Together, these accounted for 75% of coding inconsistencies.

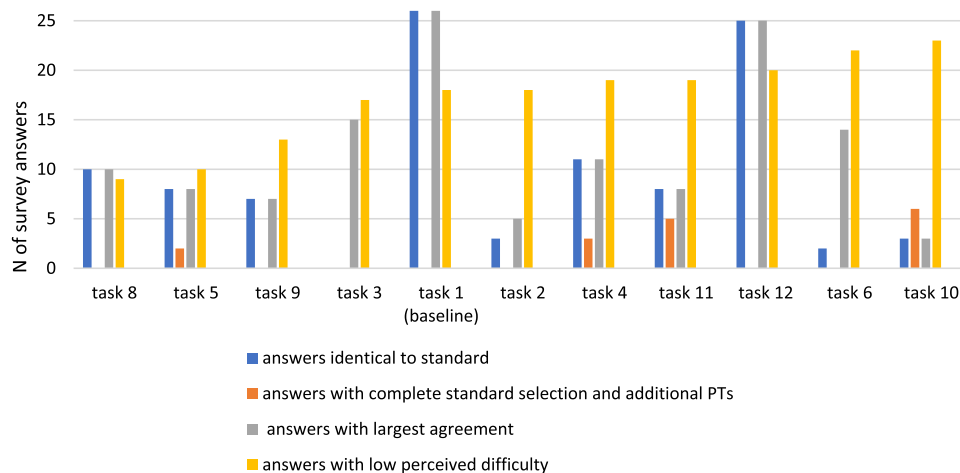
As an indication of the potential impact of each substitution in subsequent data mining of the registry, we report all

**Table 3** Demographic overview of the survey participants

Workplace	n/26 (%)
NIPH	5 (19)
NOMA	6 (23)
RELIS	15 (58)
Educational background	n/25 (%)
Medicine	4 (16)
Pharmacy	20 (80)
Other	1 (4)
Highest level of education	n/26 (%)
Bachelor	1 (3.8)
Master	21 (80.8)
PhD	4 (15.4)
Experience with MedDRA®	n/26 (%)
< 1 year	3 (11.5)
1–5 years	11 (42.3)
> 5 years	12 (46.2)
Mode of MedDRA® training (multiple selection allowed)	n/26 (%)
From colleagues	23 (88.5)
Webinars	10 (61.5)
Face-to-face MSSO course	3 (35)
Online training	6 (23)
Type of processed reports	n/26 (%)
From HCP	20 (76.9)
From non-HCP	6 (23.1)

One of the 26 survey participants did not indicate their educational background

*HCP* healthcare practitioner, *MSSO* MedDRA® Maintenance and Support Services Organization, *NIPH* Norwegian Institute of Public Health, *NOMA* Norwegian Medical Product Agency, *RELIS* Regional Medicines Information and Pharmacovigilance Centers



**Fig. 2** Clustered column chart of the number of answers identical to the standard preferred term (PT) selection (blue), the number of answers containing a complete standard PT selection and one or more additional PT(s) (orange), the maximum number of identical answers (grey), and the number of answers in which the participants evaluated

the task at a difficulty level 1 or 2 in a scale from 1 (easy) to 4 (very difficult) (yellow). The task sequence is in descending order of percentage of perceived difficulty. Note that, in tasks 1, 4, 5, 8, 9, 10, 11, and 12, the answers with the largest agreement corresponded to the standard PT selection

**Table 4** Survey tasks grouped per number of preferred terms (PTs) in the standard selection

	Task	Complete standard PT selection, percent (N)	Omission of standard PTs, percent (N)		
			No substitutions	Standard PTs + substitutions	Only substitutions
One PT in standard selection	1	100 (26)	–	–	–
	4	53.8 (14)	–	–	46.2 (12)
	12	96.2 (25)	–	–	3.8 (1)
Three PTs in standard selection	3	0	7.7 (2)	88 (22)	7.7 (2)
	5	38.5 (10)	7.7 (2)	53.8 (14)	–
	9	27.0 (7)	23.0 (6)	38.5 (10)	11.5 (3)
	10	34.6 (9)	3.8 (1)	30.8 (8)	30.8 (8)
Four PTs in standard selection	6	7.7 (2)	88.4 (23)	3.8 (1)	–
	8	38.5 (10)	11.5 (3)	46.2(12)	3.8 (1)
	11	50.0 (13)	7.7 (2)	38.5 (10)	3.8 (1)
Five PTs in standard selection	2	11.5 (3)	53.8 (14)	34.6 (9)	–

For each coding task, the table shows the percentage and number of answers containing a complete standard PT selection, as well as the percentage and number of answers in which standard PTs were omitted and/or partially or totally substituted

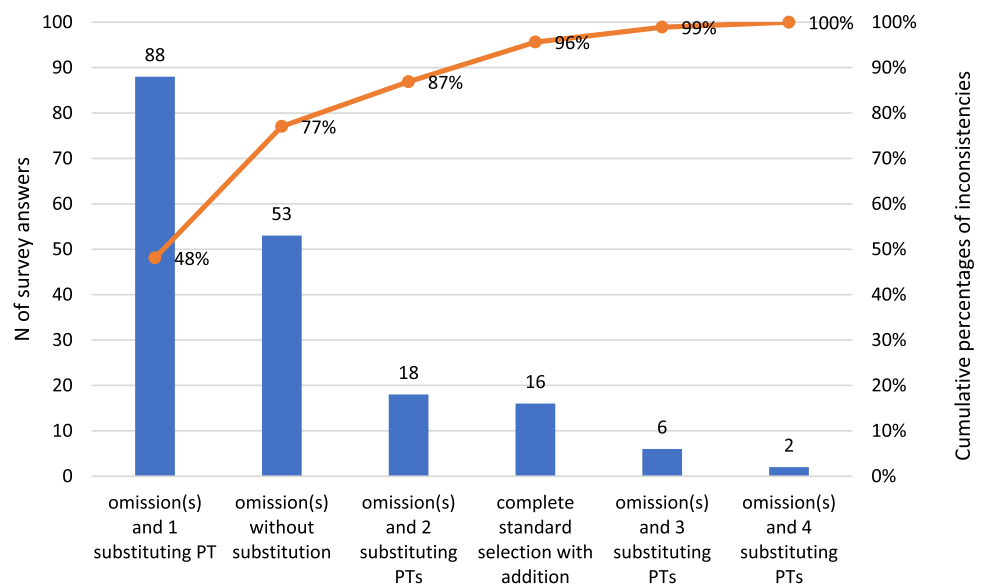
the substitutions with respective classifications into three levels of impact (Table 5). Overall, of the 32 unique PTs in the standard PT selection, 21 (65.6%) were substituted at least once, resulting in 48 unique substitutions and a total of 152 substituted PTs. Of these, 100 (65.8%) were substituted with a PT that led to the same common HLT (i.e. “Hypotension” instead of “procedural hypotension”), 36 (23.7%) were substituted by a PT that led to a different HLT but to a common SOC (i.e. “teratogenicity” instead of “maternal drugs affecting fetus”), and 16 (10.5%) were substituted by a PT that led to a different SOC. Examples of substitutions leading to different SOC were coding for diagnoses (i.e. “thyroiditis” instead of “hypometabolism,” “angioedema” instead of “peripheral swelling,” “thrombocytopenia”

instead of “platelet count decreased”), coding for investigations (i.e. “prenatal screening test abnormal” instead of “foetal cardiac disorder”), or coding for different information (i.e. “aborted pregnancy” instead of “abortion induced”). The maximum variability in substituting terms was six (see Table 5, “procedural hypotension” and “fetal cardiac disorder”).

### 3.2 Focus Group Results

Eight participants (31%) took part in the focus group interviews (Table 6). For a summary of themes, subthemes, and representative citations, see Table 3 in the ESM. Here, we summarize the content of each theme.

**Fig. 3** Pareto chart illustrating the distribution of the 183 survey answers deviating from the standard preferred term (PT) selection (bars) and cumulative impact of each type of coding inconsistencies identified (line)



**Table 5** The table shows the standard PTs that were substituted in the survey answers (column 1) and relative substituting PTs (column 2)

Standard PT, percent (N)	Alternative PT, percent (N)	Level of deviation
Asthenia, 19 (73)	Listless 2 (7.7)	PT (level 1)
Mood swings, 13 (50)	Mood altered, 5 (19.2)	HLT (level 2)
	Irritability, 1 (3.8)	HLT (level 2)
	Affective disorder, 1 (3.8)	HLT (level 2)
Hypometabolism, 0 (0)	Hypothyroidism, 20 (76.9)	PT (level 1)
	Decompensated hypothyroidism, 1 (3.8)	PT (level 1)
	Thyroiditis, 1 (3.8)	SOC (level 3)
	Thyroxine decreased, 1 (3.8)	SOC (level 3)
	Glycoprotein metabolism disorder, 1 (3.8)	SOC (level 3)
Procedural hypotension, 14 (53.8)	Hypotension, 10 (38.4)	PT (level 1)
	Hypotension + post-procedural complication, 1 (3.8)	PT (level 1)
	Hypotension + anesthetic complication vascular, 1 (3.8)	PT (level 1), HLT (level 2)
	Infusion related reaction, 2 (7.7)	PT (level 1)
	Post-procedural hypotension, 1 (3.8)	PT (level 1)
	Hypotension + anesthetic complication, 2 (7.7)	PT (level 1), HLT (level 2)
Abdominal pain lower, 12 (46.1)	Abdominal pain, 10 (38.1)	PT (level 1)
	Abdominal pain upper, 2 (7.7)	PT (level 1)
Complication associated with device, 22 (84.6)	Complication of device insertion, 1 (3.8)	PT (level 1)
	Medical device site pain, 1 (3.8)	PT (level 1)
	Product quality issue, 1 (3.8)	SOC (level 3)
	Implantation complication, 1 (3.8)	SOC (level 3)
Cerebral hemorrhage + head injury, 3 (11.5)	Traumatic intracranial hemorrhage, 1 (3.8)	HLT (level 2)
Exposure during pregnancy, 21 (80.7)	Maternal drug affecting fetus, 3 (11.5)	HLT (level 2)
	Teratogenicity, 1 (3.8)	HLT (level 2)
Exposure before pregnancy, 13 (50)	Contraindicated product administered, 1 (3.8)	HLT (level 2)
Motor developmental delay, 19 (73)	Movement disorder, 1 (3.8)	PT (level 1)
	Psychomotor skills impaired, 2 (7.7)	PT (level 1)
Speech disorder developmental, 16 (61.5)	Developmental delay, 10 (38.5)	HLT (level 2)
	Disturbance in social behavior, 1 (3.8)	HLT (level 2)
	Social (pragmatic) communication disorder, 1 (3.8)	HLT (level 2)
Fetal cardiac disorder, 13 (50)	Fetal malformation, 1 (3.8)	PT (level 1)
	Cardiac disorder, 1 (3.8)	PT (level 1)
	Congenital cardiovascular anomaly, 7 (27)	HLT (level 2)
	Cardiac failure, 2 (7.2)	HLT (level 2)
	Prenatal screening test abnormal, 2 (7.7)	SOC (level 3)
	Fetal cardiac function test abnormal, 1 (3.8)	SOC (level 3)
Abortion induced, 19 (73)	Aborted pregnancy, 3 (11.5)	SOC (level 3)
	Abortion late, 2 (7.7)	SOC (level 3)
Periorbital swelling, 11 (42.3)	Periorbital oedema, 5 (19.2)	PT (level 1)
	Eye swelling, 1 (3.8)	PT (level 1)
Peripheral swelling, 17 (65.4)	Oedema peripheral, 6 (23.1)	PT (level 1)
	Oedema, 7 (27)	PT (level 1)
	Generalized edema, 4 (15.4)	PT (level 1)
	Fluid retention, 1 (3.8)	PT (level 1)
	Angioedema, 1 (3.8)	SOC (level 3)
Swelling face, 13 (50)	Face edema, 6 (23.1)	PT (level 1)
INR increased + AST increased + ALT increased, 13 (50)	Liver function test abnormal/increased/transaminase increased/ hepatic enzyme increased, 11 (42.3)	PT (level 1)
Platelet count increased, 24 (92)	Thrombocytopenia, 2 (7.7)	SOC (level 3)

For each standard PT and substituting PT the frequency of selection in the survey answers is also indicated. For each combination standard PT/ substituting PT, the level of substitution is indicated: level 1 for substitutions leading to divergent PTs, level 2 for substitutions leading to divergent HLTs, level 3 for substitution leading to divergent SOC (column 3): PT: preferred term. HLT: high level term. SOC: system organ class  
*ALT* alanine aminotransferase, *AST* aspartate aminotransferase, *HLT* high-level term, *INR* international normalized ratio, *PT* preferred term, *SOC* system organ class

### 3.2.1 Theme 1: Aspects Related to Information Processing

Focus group participants referred to several aspects, or elements, needed to process the information described in the verbatim and classify it using MedDRA®.

One aspect was that when information was long and complex, the coder needed to determine its essence in order to assign the most appropriate MedDRA® term. This was described as challenging and sometimes required selection of multiple terms. Participants noted that, in their normal practice, it is common to use free narrative to include information that was not captured with MedDRA®.

A related aspect of information processing was categorizing the verbatims as “adverse event”, “underlying condition”, “indication”, or even “product” with the use of a drug dictionary (such as for the information on alcohol consumption, migraine prophylaxis, or intrauterine contraception). In their discussions, participants reasoned around the possible categorization in parallel with reasoning around the most appropriate term to select. They referred to categorization as an integral part of the coding task, as it helped with “sorting out” and identifying the information that was most relevant to describe the events. Sometimes, when trying to clarify the chain of events, and to distinguish an event from its consequences, participants tried to come to an agreement about how different pieces of information should be classified. When the correct classification was unclear, coders used individual evaluation. Also in this case, the strategies of multiple classifications (information can be categorized in multiple categories) and unstructured text (information that is difficult to categorize can be added as unstructured text) could be used, according to participants.

A further aspect of information processing was that lay language, common in patient reports, needed to be correctly interpreted and classified using MedDRA® terms. The perceived difficulties here were the lack of a systematic strategy and a solution that can be considered the preferred one. However, participants pointed out that not only lay language but also medical language needed to be interpreted, especially because translation from Norwegian to English is necessary. LLTs were described as very similar to each other, so finding a match in the Norwegian verbatim was potentially challenging, especially for events that the coder encountered

infrequently. Dermatological symptoms were unanimously reported as particularly difficult to code.

Finally, language is processed via culture-specific interpretation. In some cases, when told that their term choice differed from the standard PT selection proposed by the authors, participants stated that they made the choice that is correct considering the common use of the term in the Norwegian language. For instance, a participant stated that “water in the body is the Norwegian way to express ‘oedema’”. Similarly, participants pointed out that the Norwegian expression “lavt stoffskifte” (literally “low metabolism”) is used specifically to indicate low thyroid function among Norwegian health professionals and patients.

### 3.2.2 Theme 2: Ambiguity Resolution Strategies

Participants adopted some strategies to address ambiguous information. When facing vague descriptions, some selected a general term to avoid over-interpretation. Another strategy was to code multiple terms, with the aim of giving more options when unsure about which of many terms matches the described information.

Opinions varied as to whether coding information reported as uncertain was correct. Some participants thought that when information was stated as “uncertain,” it should not be coded. Others argued that, since the information they dealt with was uncertain by nature, all information should be coded and, if necessary, amended. Their main concern was missing important information.

### 3.2.3 Theme 3: Contextual Thinking and Causal Reasoning

Coders reported that their duties included causality assessment between the adverse events and suspected medicinal product. Some indicated that they bore this in mind throughout the selection of the MedDRA® term, using it to guide what they deemed to be the most appropriate selection. Some participants also regarded information deemed relevant to causality assessment as justifiably eligible for coding, irrespective of its certainty; for instance, if it could provide a possible alternative explanation for the described symptom.

Some of the participants expressed that the coding tasks were sometimes more difficult than real-life tasks because the suspected medicine was unknown. Some agreed that knowing the suspected medicine(s) could guide the choice of whether to code uncertain information. In particular, some accounted for the pharmacological mechanism when determining which information to code and which to omit. For example, in task 6, a centrally acting medicinal product, but not an anticoagulant, would have led some to code “head injury”. Others disagreed and insisted that coding should have been the same, irrespective of the suspected medicine. In addition to the lack of suspected medicine, participants

**Table 6** Description of the focus groups

Focus groups	Participants (n)	Represented institutions (n)	> 5 years' experience (n)	2–5 years' experience (n)	< 2 years' experience (n)
1	4	2	2	2	0
2	2	1	1	0	1
3	2	2	1	1	0

thought that, in general, lack of contextual information undermined contextual thinking, making it difficult to select the best-fitting MedDRA<sup>®</sup> term.

### 3.2.4 Theme 4: External Resources

Participants named several resources they used when they lacked contextual information. The most immediate and low-threshold daily practice was consultation with peers. Outside the survey, consultations with medical doctors occurred when coders were unfamiliar with verbatims or when the verbatims did not match with MedDRA<sup>®</sup> terms. In these cases, participants also reported using Google searches. Another point of contact for coders could have been the original reporters when the verbatims were insufficient or vague. Such contacts were described as contingent on a case-by-case evaluation according to the seriousness of the reported event(s), time constraints, and overall report quality. Some noted that, when the report was very poorly described, contacting the reporter did not yield better information. Participants also spoke of MedDRA<sup>®</sup> resources that were routinely used for MedDRA<sup>®</sup> term selection. For instance, informants actively used the MedDRA<sup>®</sup> hierarchy and assessed which concept the term belonged to in the higher levels of the hierarchy. In addition, all participants said they were familiar with the MTS:PTC guideline. However, they reported performing situational evaluations on the applicability of guidelines to certain ambiguous verbatims. For instance, one participant who was aware that item 3.2 of the MTS:PTC prescribes the coding of provisional diagnoses declared that it could not apply to extremely uncertain tasks, such as N6.

### 3.2.5 Theme 5: Wish for Systematic Training that Includes an Overview of Pharmacovigilance Activities Downstream Coding

Participants reported that MedDRA<sup>®</sup> training was unsystematic. Some felt that they started the job with insufficient training and were eventually introduced to a structured MedDRA<sup>®</sup> Maintenance and Support Services Organization course. All participants considered such a course as very useful, and some reported having gained a whole new understanding of the coding task after the course. Participants wished for a more systematic and periodical training offer to refresh and update their knowledge. They reported that adequate training would stress the importance of coding to the whole downstream process of signal detection and analysis of signals. With this awareness, the coder would be motivated to capture the relevant information as accurately as possible, rather than seeing coding as a mere bureaucratic duty. Some participants proposed that a way to foster the role of the coder within the whole pharmacovigilance process

would be to collaborate with other institutions for different tasks. For instance, during the COVID-19 vaccination campaign, meetings were held between coders and assessors to discuss signals, and collaborative coding efforts were undertaken for large numbers of reports.

## 3.3 Integrated Results

By summarizing and integrating survey and interview data, we highlight the main findings indicated in Table 7.

## 4 Discussion

To our knowledge, this is the first qualitative study exploring MedDRA<sup>®</sup> coding practices. Other studies in the literature provide a statistical perspective of the level of consistency across different MedDRA<sup>®</sup>-relevant platforms. To improve coding inconsistency, we need to better understand its root causes. Our analysis offers an initial step in this direction. In line with the study aim, we used purposive sampling to identify the selection of coding tasks. This type of sampling allowed us to study a specific phenomenon, in our case ambiguity. Under the assumption that ambiguous information is the most difficult to code, our purposeful sampling was more likely to generate high levels of inconsistency, which do not statistically represent Norwegian coding practices. Instead, we explored the possible grounds for such inconsistency. A strength of this study was the user participation and that the research problem was suggested by a central member of the user group and co-author (HS). The large participation in the survey phase indicates the relevance of the research problem to Norwegian coders. Another advantage was the study's mixed-methods design, in which participants solved the task in written form and were then provided the opportunity to explain and comment on their coding choice. We suggest that our survey, followed by group discussion, can be used by single agencies to map internal MedDRA<sup>®</sup> coding practices for capacity building.

### 4.1 Main Findings

We found cases in which survey participants had a large percentage of agreement on a coding choice that was inconsistent with the authors' choice of standard PT selection (tasks 3 and 6, Fig. 2). One of the explanations identified in the focus groups was linked to culture-specific connotations of medical language. Norwegian survey participants were confident that they had correctly interpreted the information in the Norwegian "verbatim," but the literal translation into English yielded a different term. In this case, the codes provided by the participants are contextually better situated and therefore likely more in line with the intended meaning

**Table 7** Integrated overview of the main findings from the two study phases

Type of inconsistency observed (survey)	Explanations (focus groups)	Example	Example quote
Inconsistency with standard PT selection, but good consistency among participants and low perceived difficulty (Fig. 1, tasks 3 and 6)	Aspects related to information processing: There is a cultural connotation of language	“Low metabolism” (Norwegian “lavt stoffskifte”) coded as “hypothyroidism” instead of “hypometabolism”	“But it is a typical Norwegian way of writing. But the concept ‘stoffskifte’ does not exist in English. So, one would code this as ‘hypothyroidism’ or something similar” (interview 1)
	Ambiguity resolution strategy: Uncertain information should not be coded	“Not sure if there was head trauma” remained uncoded	“I didn’t include a head injury. It was uncertain so at least I’m clear that I shouldn’t include that” (interview 2)
Substituting one out of multiple PTs (the most common type of inconsistency, Table 4 and Fig. 3)	Aspects related to information processing: Finding the medical word for lay language	“feeling moody” either as “mood swings, mood altered” or “irritability”	“First one must understand what the reporter has meant in Norwegian or interpret it correctly. It’s not always that Norwegian medical terms are used either. If it says high blood pressure or hypertension in Norwegian, then it’s easy. But if it says the patient feels groggy, then it’s a bit more difficult” (interview 2)
	Aspects related to information processing: Finding the right medical term in English	“motoric delay” as “motor developmental delay, movement disorder” or “psychomotor skills impaired”	“... it’s about finding the right English term, but also there are so many terms that are almost the same, or that can be a bit more precise, exactly what is meant, than a more general term. So, we have to use our imagination and search a bit in the MedDRA® browser, also on medical terms quite often” (interview 1)
	Aspects related to information processing: Fitting a whole complex description in MedDRA® terms	“hypotension” combined with “post procedural hypotension” as substitutes for “procedural hypotension”	“And sometimes there isn’t a term that covers the entire event, but you have to code several terms to describe the problem from slightly different angles, depending on what type of problem it is, and what it is something around cause” (interview 1)

Table 7 (continued)

Type of inconsistency observed (survey)	Explanations (focus groups)	Example	Example quote
Omitted terms without substitution (Table 4 and Fig. 3)	<p>Ambiguity resolution strategy: Uncertain information should not be coded</p> <p>Contextual thinking and causal reasoning: Knowledge of the causality assessment process influences which information is considered relevant</p> <p>Aspects related to information processing: categorizing the information</p> <p>Contextual thinking and causal reasoning: Considering the characteristics of reported medicine (partial agreement)</p>	<p>“resembles skin cancer” remains uncoded</p> <p>Choice on whether the information “patient using alcohol along with the medication” should be coded as event-relevant</p> <p>Choice on whether the information “patient using alcohol along with the medication” should be coded as event-relevant</p> <p>Choice on whether to code the information “Not sure if there was head trauma”</p>	<p>“A spot on the forehead that resembles skin cancer, then it's kind of ... it's not something that kind of proves that it's skin cancer. Yes, I don't really know. I probably wouldn't have taken skin cancer” (interview 3)</p> <p>“It is an alternative explanation of what has happened. That is, the patient's elevated liver values and this. Alcohol use can also cause liver disease. But I would include it in my assessment when I was going to do the causality assessment, and I coded it. Because there are other possible explanations than just this tablet supplement” (interview 1)</p> <p>“I don't know if it was wrong to code it in as a side effect, because it's not ruled out. I don't know if it belongs more under product or possibly disease. 'Alcohol use' could be placed under there.” (interview 3)</p> <p>“But I think it's very dependent on which medication is suspected. If it's a medication that affects the central nervous system, or has anticholinergic effects, and can cause dizziness and hypotension, and those types of things that make the patient fall more often, then I might think about it [to include head injury]. Because if the patient falls because of side effects of the drug, and it causes a head injury, then it's quite serious. If it was an anticoagulant, then I don't think it causes a head injury. I think I would have put it on concomitant disease, and written something about it being uncertain whether it actually occurs” (interview 1)</p> <p>“It's reported on suspicion so the coding should be the same [regardless of which medicine it concerns]” (interview 3)</p> <p>“Some of those skin reaction cases just end up in a non-specific bag called rash. Because it's a type of rash. But we can't find a specific sign” (interview 1)</p> <p>“I think that the swelling is most likely edema, but you can't quite say it. That's why I chose both edema and swelling” (interview 3)</p>
	<p>Ambiguity resolution strategy: coding vague description by using unspecific terms</p> <p>Ambiguity resolution strategy: Coding vague descriptions by using multiple terms</p>	<p>“hepatic enzyme decreased” to express to code the information about three enzymes</p> <p>“water in the body” coded as “swelling” combined with “edema”</p>	
Addition (Table 3 and Fig. 2)			

Table 7 (continued)

Type of inconsistency observed (survey)	Explanations (focus groups)	Example	Example quote
Substitutions that generate an inconsistency at the SOC level	Aspects related to information processing: categorizing the information  Insufficient training	<p>“during pregnancy screening a serious heart failure is observed” coded either as investigations (“prenatal screening test abnormal”) or symptom (“fetal cardiac disorder”)</p> <p>Errors: coding for diagnoses (i.e. “thyroiditis” instead of “hypometabolism, angioedema” instead of “peripheral swelling,” “thrombocytopenia” instead of “platelet count decreased”); coding for a different information than in the verbatim (i.e. “aborted pregnancy” instead of “abortion induced”)</p>	<p>“And then cardiac malformation confirmed as a test result. It should not be as a side effect” (interview 1)</p> <p>“I have forgotten how I have done [my MedDRA® learning process], because it wasn't so structured with MedDRA®. It was a bit like, here's MedDRA®, it works like that, here you are” (interview 1)</p>

of the original statement. MedDRA® training resources emphasize that “the MTS:PTC document does not address every potential term selection situation. Medical judgment and common sense should also be applied” [7]. Our results indicate that organizations might sometimes need to apply judgement and agree on local preferred coding options. The need to adjust international guidelines locally, for instance by user involvement and contextual thinking, has been identified in the broad medical literature [14–17], and our results indicate that MedDRA® is no exception.

A second finding was that the common knowledge of MedDRA® guidelines was sometimes insufficient to align coding decisions (see task 6). MTS:PTC recommends that no information should be added and that all information should be coded [8]. When information is ambiguous, the coder might be unsure whether a “head injury” has in fact occurred. As such, if “head injury” had occurred, the coder would make an omission by not coding it. However, if a “head injury” had not occurred, the coder would make an addition by coding it. Inductive risk theories address these kinds of situations, in which the set of possible decisions wholly includes imperfect outcomes [18, 19]. Inductive risk, stating that every choice needs a formal answer to the question on which type of potential error is preferable, has been applied to most fields of decision-making in science [20] and suggests the application of such theories in future research. Currently, the responsibility to resolve ambiguity is on the coders but may have to be shared between the coders and organizations through standard coding practices and procedures. If standard procedures are to align the coding choices, we suggest that they should state a preferred strategy for ambiguity resolution. If, at an institutional level, it is preferred to avoid additions in such cases, there will be a level of missed information. If, on the other hand, it is preferred to avoid omissions in such cases, there will be a level of over-represented information.

A third finding of this study is that placing the information in the perspective of causality assessment can influence the coder's choice to code or dismiss information. In Norway, this can happen since some pharmacovigilance officers are in charge of coding and the subsequent causality assessment of the case. MedDRA® training states that causality is not related to MedDRA® coding. However, it might be difficult for some coders to set aside the background causal thinking when they are in charge of both duties. Future research should verify whether coding and causality assessment are overseen by the same staff in other settings and investigate the effects on coding of this practice.

Finally, our survey detected a few inconsistencies that could be classified as errors because they failed to adhere to a specific MedDRA® recommendation, for instance, coding for a diagnosis when the information only contains symptoms (Table 5). We speculate that this might be due to insufficient

training, in line with statements from some focus group participants who wished for more systematic training. However, the qualitative nature of our sampling strategy meant we could not statistically test whether training experience predicts these types of errors. Future research with a statistical sampling of the registry should address this question to clarify the effect of current MedDRA<sup>®</sup> training forms.

For the first time, we have identified and described specific challenges associated with coding ambiguous information using MedDRA<sup>®</sup>. In light of our findings, it becomes even more urgent to address the prevalence of ambiguous verbatims in databases of individual case safety reports, especially given the increasing contribution of patient reporting. Existing research in this area is sparse, focuses on coders trained for HIV clinical trial events, and does not encompass the complexity of reporting [10].

## 4.2 Study Limitations

This study has some limitations. The majority (69%) of survey participants withdrew from the focus groups, so we do not know whether we reached data saturation. However, our intention was not to provide a full list of all types of inconsistencies that can occur. Given the richness of the focus group data, recurring patterns across the thematic analysis, and matching of survey observations with qualitative themes, we consider that we have answered the research question.

Another limitation is that a concept validity issue with the survey was detected during the qualitative study phase. Participants might have interpreted the survey task “coding information about the adverse event” as indicating only the LLTs that in real practice they would fill in the structured space dedicated to “event.” For instance, some participants did not code “migraine prophylaxis”, not because they did not find the information important, but because in real practice the information about the suspected medicine would have been coded as “product” with another dictionary, and the information on migraine as “indication.” Therefore, the risk existed of inflating the counts of inconsistencies in coding due to factors external to our research question. However, detection of this issue allowed us to exclude controversial terms. For instance, “alcohol use” was discussed in the focus groups as potentially event related but excluded from the quantitative analysis since some participants stated that it belonged to “medical history,” “products,” or others. Therefore, our survey analyses indicate the minimum amount of coding variation. It is possible that the phenomenon is larger than we have defined here.

Furthermore, we tried to indicate the potential consequences of coding inconsistencies using a hierarchy analysis. However, the real impact of the inconsistencies on signal detection is difficult to establish from these three levels and

depends both on the data retrieval strategy (via SOC, standardized MedDRA<sup>®</sup> query, or both) and on the analysis level (data mining/clinical assessment). Indeed, for the clinical evaluation of a case series, a PT-level substitution may play a big role if the information in the original report is not accessible. For instance, “abdominal pain upper” and “abdominal pain lower” belong to the same HLT but can be symptoms of two very distinct conditions.

Finally, the survey length was constrained by feasibility requirements, allowing only a limited number of questions to ensure adequate recruitment and completion rates. A further study might better characterize the phenomenon by designing multiple surveys, each focusing on one of the sources of inconsistency identified here.

## 5 Conclusion

This study explored coding consistency among coders and between coders and a MedDRA<sup>®</sup> standard PT selection and examined the types of inconsistencies and the thought processes behind them. Our findings indicate that consistency levels varied significantly depending on the specific coding task. In our selection of tasks, this heterogeneity was influenced by the type of information in the verbatim rather than the number of verbatims.

The most common inconsistency arose from substituting one of multiple terms. Themes from focus groups explaining these substitutions included difficulties in translating lay language to medical terminology, finding accurate English translations for Norwegian medical terms, and fitting complex descriptions into MedDRA<sup>®</sup> terms. Some substitutions resulted in inconsistencies at the SOC level, partially due to the choice between investigation terms and terms for symptoms or diagnoses. Other SOC-level inconsistencies could not be explained and may be linked to insufficient training, although this theme was not associated with any particular task.

A second type of inconsistency involved omitting terms without substitution. Themes explaining these omissions included strategies for resolving ambiguity, contextual thinking, causal and pharmacological reasoning in the coding process, and information categorization. A third type of inconsistency was the unnecessary use of multiple terms, which was explained by themes related to ambiguity resolution strategies.

Lastly, the variation in coding performance did not align with the perceived difficulty of the tasks. For instance, some coding tasks exhibited high inconsistency with the standard PT selection but demonstrated good consistency among participants and low perceived difficulty. These cases were partially explained by the theme of cultural connotations of language.

Clearer institutional guidelines and routines, as well as tailored training programmes need to be developed to target the sources of coding inconsistencies suggested in this study. Future research is needed to quantify the impact of the challenges we identified on the coding accuracy of ambiguous information in pharmacovigilance registries.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40264-025-01573-2>.

**Acknowledgements** The authors are grateful to Jane Knight, Clinical Associate with the MedDRA Maintenance and Support Services Organization, for help and guidance with the purposive sampling of ambiguous coding tasks and consultancy regarding the standard term selection. The MedDRA® trademark is registered by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use.

## Declarations

**Funding** Not applicable.

**Conflict of interest** All authors have no relevant conflicts of interest to declare. The views expressed are those of DS and not necessarily those of the Uppsala Monitoring Centre.

**Ethics approval** This research was carried out in accordance with relevant guidelines and regulations. The regional research ethics committee south-east Norway concluded that the project fell outside the scope of the Norwegian Health Research Act (reference number 539798, 17.04.2023). Data collection and storage were designed in consultation with the Norwegian Agency for Shared Services in Education and Research to ensure adherence to data protection and privacy regulations (reference number 766729, 10.05.2023).

**Consent to participate** Focus group participants provided written informed consent. The survey was conducted anonymously after participants had received an information letter.

**Consent for publication** All participants consented to the publication of the research data.

**Availability of data material** Personal data are not openly available for reasons of sensitivity. De-identified raw research data are available from the corresponding author upon reasonable request.

**Code availability** Not applicable.

**Author contributions** ER defined the aim of the study, conceived the study design, was responsible for consultation with the Norwegian Agency for Shared Services in Education and Research to ensure adherence to data protection and privacy regulations, conducted the final data analysis in consultation with the other authors, and drafted the manuscript. TG carried out the purposive sampling, drafted the survey and the interview guide, led the group interviews, and conducted a preliminary data analysis. HS proposed the aim of the study, led the purposive sampling, led the recruitment of participants, and consulted on the final version of the survey and interview guide. DS consulted on the final versions of the survey and interview guide and provided support in defining the study design and data analysis. FA consulted on the qualitative data analysis and helped with the drafting of the manuscript. DJ consulted on quantitative data analysis and conceived Table 4. All

authors gave feedback and proposed changes to the first manuscript draft. All authors read and approved the final version.

## References

1. Mugosa S, Stankovic M, Turkovic N, et al. Medical dictionary MedDRA: used in over 60 countries, among which is Montenegro. *Hosp Pharmacol*. 2015;2:266–71.
2. Lindquist M. VigiBase, the WHO Global ICSR Database System: basic facts. *Drug Inf J*. 2008;42:409–19.
3. Schroll JB, Maund E, Gøtzsche PC. Challenges in coding adverse events in clinical trials: a systematic review. *PLoS ONE*. 2012;7:e41174.
4. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). *Drug Saf*. 1999;20(2):109–17.
5. Mozzicato P. MedDRA: an overview of the Medical Dictionary for Regulatory Activities. *Pharm Med*. 2009;23(2):65–75.
6. Brown EG. Using MedDRA: implications for risk management. *Drug Saf*. 2004;27(8):591–602.
7. MedDRA Maintenance and Support Services Organization. MedDRA coding basics. 2024. [https://files.meddra.org/www/Training%20Materials/2024/Materials/001042\\_MedDRA\\_Coding\\_Basics\\_Webinar.pdf](https://files.meddra.org/www/Training%20Materials/2024/Materials/001042_MedDRA_Coding_Basics_Webinar.pdf).
8. MedDRA Maintenance and Support Services Organization. Support documentation. <https://www.meddra.org/how-to-use/support-documentation/english>.
9. Combi C, Zorzi M, Pozzani G, et al. From narrative descriptions to MedDRA: automatically encoding adverse drug reactions. *J Biomed Inform*. 2018;84:184–99.
10. Tonéatti C, Saïdi Y, Meiffredy V, et al. Experience using MedDRA for global events coding in HIV clinical trials. *Contemp Clin Trials*. 2006;27:13–22.
11. Brajovic S, Blaser DA, Zisk M, et al. Validating a framework for coding patient-reported health information to the Medical Dictionary for Regulatory Activities Terminology: an evaluative study. *JMIR Med Inform*. 2018;6(3):e42.
12. Kralova K, Wilson C-A, Richebourg N, et al. Quality of MedDRA® coding in a sample of COVID-19 vaccine medication error data. *Drug Saf*. 2023;46:501–7.
13. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3:77–101.
14. Anjum RL, Copeland S, Kerry R, et al. The guidelines challenge—philosophy, practice, policy. *J Eval Clin Pract*. 2018;24:1120–6.
15. Solomon M. Making medical knowledge. Oxford: OUP; 2015.
16. Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? *BMJ*. 2014;348:g3725–g3725.
17. Greenhalgh T, Fisman D, Cane DJ, et al. Adapt or die: how the pandemic made the shift from EBM to EBM+ more urgent. *BMJ Evid Based Med*. 2022;27:253–60.
18. Douglas H. Inductive risk and values in science. *Philos Sci*. 2000;67:559.
19. Andersen F, Rocca E. Underdetermination and evidence-based policy. *Stud Hist Philos Biol Biomed Sci*. 2020;84: 101335. <https://doi.org/10.1016/j.shpsc.2020.101335>.
20. Elliott KC, Richards T, editors. Exploring inductive risk. Oxford: OUP; 2017.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Authors and Affiliations

Tahmineh Garmann<sup>1</sup> · Hilde Samdal<sup>2</sup> · Daniele Sartori<sup>3</sup>  · David Jahanlu<sup>1</sup> · Fredrik Andersen<sup>4</sup> · Elena Rocca<sup>1</sup> 

✉ Elena Rocca  
elena.rocca@oslomet.no

<sup>1</sup> Faculty of Health Sciences, Institute for Natural Science and Health, Oslo Metropolitan University, Postboks 4, St.Olavs plass, 0130 Oslo, Norway

<sup>2</sup> Norwegian Medical Product Agency, Oslo, Norway

<sup>3</sup> Uppsala Monitoring Centre, Uppsala, Sweden

<sup>4</sup> Faculty of Health, Welfare and Organization, Østfold University College, Fredrikstad, Norway