An Exploration of Risk in the Use of MT in Healthcare Settings with Abbreviations as a Use Case

Eleanor Taylor-Stilgoe, Constantin Orăsan\[0000-0003-2067-8890\] and Félix do Carmo\[0000-0003-4193-3854\]

University of Surrey, Guildford, Surrey, GU2 7XH, United Kingdom
{e.j.taylor-stilgoe,c.orasan,f.docarmo}@surrey.ac.uk

Abstract. When faced with language barriers, UK healthcare staff have found themselves turning to machine translation (MT) – predominantly Google Translate – to fulfill their duty of care to patients [3, 16, 23]. Despite the risks potentially posed by the use of MT in such complex and sensitive situations, little research currently exists as to healthcare staff awareness of these risks in real-life settings. This gap is particularly notable concerning the use of MT with patient medical record information compared with interpersonal situations and patient-oriented documentation [6, 7, 8, 14, 16, 22, 27]. While research has been conducted into the perceptions and practices of the general population concerning MT use in largely lower-stakes contexts [31], research on the extent to which these transfer to higher-stakes settings remains lacking. The contribution this paper aims to make is therefore twofold: to investigate the impact of MT on patient medical record documentation and to explore the extent to which healthcare staff are aware of the risks potentially posed by its use. In this paper, we selected contextualised medical abbreviation examples from authoritative French and Spanish clinical corpora [9, 15] to serve as a use case, abbreviations having previously been shown to pose an increased risk for patient harm even prior to their translation with Google Translate [2, 5, 12, 19, 25, 28]. Examples containing higher-risk MT errors were presented to healthcare staff to ascertain their perceptions and risk awareness as part of semi-structured interviews. Whilst these interviews remain ongoing, this paper presents the findings on risks identified in the use of MT with patient medical documentation, and the responses obtained thus far.

Keywords: machine translation, Google Translate, healthcare, patient medical records, medical abbreviations, risk awareness

1 Introduction

The National Health Service (NHS) in England formally cautions its staff against using online MT services on the grounds that ‘there is no assurance of the quality of the translations’ [17]. Despite this, cases abound in which healthcare staff have found themselves resorting to non-domain specific, commercially available MT, such as Google Translate (GT), when providing interpersonal or written assistance to patients with limited to no English language proficiency [3, 16, 23]. Vieira et al. [30] note that ‘research on the implications of the widespread and potentially uninformed use of this technology remains sparse’. Further to this, despite patient medical records constituting an arguably central component of patient care, there is currently next to no literature on the implications of their
use with MT. The need for further research in this specific yet significant subdomain is underscored by the comparative attention paid to the use of MT as a form of interpreting aid in interpersonal situations and for patient-oriented documentation [6, 7, 8, 14, 16, 22, 27]. The contribution this paper therefore aims to provide is an investigation of the impact of MT on patient medical documentation, and an examination of the extent to which healthcare staff are aware of the risks potentially posed by its use.

To this end, contextualised examples containing medical abbreviations were selected as a use case, these ubiquitous features of the healthcare domain having been shown to pose an increased risk for patient harm even at the monolingual level [2, 5, 12, 19, 25, 28]. This paper presents an analysis of the impact of GT on specific abbreviation examples in terms of the errors produced at a linguistic level and the potential risks to patient safety arising from its use. Examples containing higher-risk MT errors were presented to healthcare staff within the context of semi-structured interviews to ascertain their perceptions of and risk awareness of the same. While interviews with healthcare staff remain ongoing, the preliminary conclusions drawn thus far are provided in Section 8 below.

2 Related Work

The misinterpretation and/or misuse of medical abbreviations in healthcare settings has been identified in the literature as a significant contributing factor towards increased potential for patient harm, even at the monolingual level [2, 5, 12, 19, 25, 28], though scant research comparatively exists with regard to their translation using MT. It is for precisely this reason that abbreviations were selected as a use case with which to analyse phenomena for which the risk to patient safety in such higher-stakes contexts may potentially be further exacerbated by the use of commercially available and non-domain specific MT.

In an attempt to overcome the difficulties posed by language barriers in healthcare settings, various domain-specific applications have been developed over the last twenty years with a view to improving delivery of care to patients from diverse cultural and linguistic backgrounds [1, 4, 18, 26]. Further to this, a range of mobile translation applications developed specifically for providing communication support to both staff and patients unable to speak the dominant language of the healthcare setting have been subject to formal evaluation as to their effectiveness in practice [11, 20, 21, 24, 29]. Despite these developments, commercial MT in general and Google Translate in particular remain prevalent in both academic and grey literature discussing the use of technologies in healthcare settings [3, 6, 7, 8, 14, 16, 22, 27]. The ready availability, convenience of use and cost-effectiveness of MT appear to make it a ‘better than nothing’ alternative for staff who must balance their duty of care against time, cost, and/or resourcing constraints [3, 16, 23]. As noted by Vieira et al. [30], ‘language barriers coupled with funding pressures and other practical difficulties expose doctors and patients to a dilemma where MT, albeit risky, is perceived as the easiest route to cross-linguistic communication’. Google Translate was therefore selected for use in the present research, reflecting as this does its documented use in real-life healthcare contexts.

Furthermore, while research has been conducted into the perceptions and practices of the general population concerning MT use in primarily lower-stakes contexts [31], studies on the extent to which these transfer to such higher-stakes settings as healthcare remain limited. This paper therefore also seeks to contribute to the knowledge gap in regard to the extent to
which healthcare staff are aware of the risks potentially posed by its use when applied to professional settings involving patient care.

Finally, the existing literature on the use of MT in healthcare settings is largely focused on its use as a form of interpreting aid in interpersonal situations and for patient-oriented documentation [6, 7, 8, 14, 16, 22, 27]. As such, this paper aims to examine its use within the context of patient medical record information, this being a particularly understudied yet significant aspect of an already understudied area.

3 Data Research Methodology

3.1 Data Collection

Medical abbreviations were selected as a use case with which to investigate and analyse the creation or amplification of risk through their use with MT. This decision was made on account of such abbreviations having previously been identified as posing an increased risk for patient harm even prior to translation [2, 5, 12, 19, 25, 28]. It should be noted that ‘abbreviation’ is, per Sheppard et al. [25], used as an umbrella term for ‘any shortened form of a word […] any acronym […] any contracture […] and any initialism’.

It was first necessary to select at least one in-domain, non-English SL corpus with which to identify and isolate instances of relevant phenomena. A key criterion was the prioritisation of reliable clinical data from authoritative sources. The aim of this was to echo as authentically as possible the medical nature and linguistic features of written information encountered in healthcare environments, allowing for a more realistic assessment of the potential risks involved in its translation with MT. The decision was made to begin with source languages with which the first author was already familiar, namely French and Spanish.

The first corpus selected was the CAS corpus, a French clinical corpus comprising clinical cases published in scientific literature and both legal and medical training resources [9]. This contained a mixture of real, de-identified and fake patients used for medical educational purposes and comprised 4,900 clinical cases. The second corpus selected was CodiEsp, a Spanish clinical corpus comprising a randomly sampled development dataset (250 cases), test set (250 cases) and training set (500 cases), along with an unannotated background set containing a further 2,751 cases [15]. A notable advantage offered by both corpora was their coverage of various specialties, patient comorbidities and case histories written at the expert-to-expert level, as would be expected from the type of real-life healthcare documentation on which both were based.

3.2 Data Preparation and Processing

Given the size of the corpora used, a simple custom desktop application with which to identify and isolate relevant data instances was designed by the first author and implemented by a professional software developer using Microsoft Visual Studio and C#. This allowed for the filtering of sentences from the corpus according to manually configurable and selectable search patterns, implemented using regular expressions. The option to manually add matches containing irrelevant instances to a list of exclusions was also incorporated, as was a feature enabling exportation of selected matches to an Excel file for use with MT and analysis in
Sketch Engine [13]. Any noise present in the data, such as spelling or grammatical errors, was preserved with the aim of recreating in the MT output those issues encountered in real-life settings. The paragraph within which each instance occurred was also preserved according to the preceding and subsequent line breaks to ensure that sufficient context was retained. The retrieved matches were then translated with GT.

4 Data Examples: Experimentation, Results, and Analysis

A random set of SL matches were analysed to explore any issues arising from their translation with GT. In addition to focusing on abbreviations, any other relevant phenomena were acknowledged and recorded for reference, with the research remaining focused on the use case. The following examples were considered particularly notable in terms of their potential for increased risk to patient safety and were thus selected for use in interviews with healthcare staff. As follows (emphasis and/or [sic] added for clarity):

Example 1: ‘AO’

ST: En el examen inicial presentaba una AV con su corrección (AV CC) de 0,2 en AO que mejoraba a 0,4 con estenopeico. La motilidad pupilar era normal, la presión intraocular (PIO) de 18mm Hg en AO y no presentaba alteraciones en la biomicroscopia anterior (BMA). La refracción bajo cicloplejia (RBC) era de -4,50 en AO. En el fondo de ojo (FO) se apreciaban estrías retinianas en la mácula de AO. La tomografía axial computarizada (TAC) no reveló anomalías en la región orbitaria ni asimetrías ni alteraciones en la posición del diafragma iridocrystaliniano.

GT: In the initial examination, she presented a VA corrected for her (VA CC) of 0.2 in AO that improved to 0.4 with pinhole. Pupillary motility was normal, intraocular pressure (IOP) was 18mmHg in AO, and she had no alterations in anterior biomicroscopy (AMB). Refraction under cycloplegia (RBC) was -4.50 in AO. Retinal striae were seen in the eye fundus (FO) in the AO macula. Computerized axial tomography (CAT) revealed no abnormalities in the orbital region or asymmetries or alterations in the position of the iridocrystalline diaphragm.

This Spanish to English ophthalmological example contains four instances in which the abbreviation ‘AO’ (‘ambos ojos’ in Spanish, for which the correction translation would be the Latin ‘OU’, ‘oculus uterque’ or English ‘BE’, ‘both eyes’) remains wholly untranslated in the MT, seemingly due to the lack of a full term preceding the SL abbreviation. The result is the omission of central clinical information as to the relevant examination having been conducted on both eyes. While other abbreviations present in the MT, such as ‘RBC’ (‘refracción bajo cicloplejia’/’cycloplegic refraction’, ‘CRx’) and ‘FO’ (‘fondo de ojo’/’[eye] fundus’) remain either mistranslated or untranslated, these are at least preceded by the full if only broadly rendered terms, whereas this surrounding context remains absent for ‘AO’.

Example 2: ‘DRS’

ST: Au cours des dix dernières années, elle a subi quatre coronarographies, neuf échographies cardiaques et douze épreuves d’effort au tapis roulant. Depuis la dernière dilatation, quatre ans auparavant, aucune lésion significative ou donnée probante de resténose ne sont notées aux
coronarographies. Les pressions pulmonaires sont normales. Aucune maladie valvulaire n’apparaît aux échographies. Les paramètres électrophysiologiques en relation avec les suivis liés à son stimulateur cardiaque défibrillateur sont adéquats[sic]; aucune anomalie significative n’apparaît et aucune thérapie ne lui est donc appliquée. Bien que cette dame ne soit pas suivie dans cet établissement, elle nous assure que son lupus est stable. D’ailleurs, la protéine C réactive est normale, la tension artérielle est bien contrôlée, et les douleurs angineuses n’ont pas un caractère péricardique. Enfin, elle n’est connue pour aucun antécédent psychiatrique. Pourtant, la patiente continue de présenter des DRS.

GT: Over the past ten years, she has had four coronary angiograms, nine cardiac ultrasounds and twelve treadmill stress tests. Since the last dilation, four years earlier, no significant lesions or evidence of restenosis have been noted on coronary angiograms. Pulmonary pressures are normal. No valvular disease appears on ultrasound. The electrophysiological parameters in relation to the follow-ups related to his defibrillator pacemaker are adequate; no significant abnormality appears and no therapy is therefore applied to him. Although this lady is not followed in this establishment, she assures us that her lupus is stable. Moreover, C-reactive protein is normal, blood pressure is well controlled, and angina pain is not pericardial in nature. Finally, she is not known to have any psychiatric history. However, the patient continues to present with DRS.

This largely cardiological French to English example, which shifts to psychiatry in the penultimate sentence, reflects a complete failure by the MT to translate ‘DRS’ (‘douleur(s) rétrosternale(s)’), for which the correct translation in English would be ‘retrosternal pain’. While this non-translation in the MT also appears to have arisen from the lack of a full term preceding the SL abbreviation, by comparison with ‘AO’ for which a directly equivalent medical abbreviation does indeed exist in either Latin or English, both options being in use in modern healthcare settings, no such direct equivalent exists in English for ‘retrosternal pain’. Despite being the only abbreviation in the paragraph, ‘DRS’ is of central importance to identifying the patient’s symptoms, described as these are as ongoing. The ambiguity arising from this non-translation may hinder the patient’s treatment as a result.

In view of the untranslated abbreviations in the MTs and the apparent centrality of these to comprehension, it was speculated that the ambiguity arising from such issues may at best hinder effective care delivery or, at worst, actively increase the potential for patient harm in real-life healthcare settings. Interviews with healthcare staff were planned to explore their views on this potential for increased risk accordingly.

5 Interview Methodology

5.1 Recruitment Process

Participants approached for recruitment included both medically qualified professionals and those performing non-clinical healthcare roles, this professional diversity being considered directly relevant to the type of information each might seek to identify in the texts, their respective degrees of clinical risk awareness, and their approaches to risk management. No specific criteria were assigned concerning participants’ foreign language knowledge, length of professional service or specific care setting, these aspects all being considered to offer both more realistic diversity in terms of background and potentially interesting analyses.
Recruitment was conducted via snowball sampling. The study was granted a Favourable Ethical Opinion by the University of Surrey Ethical Committee in February 2023.

5.2 Interviews: Data Example Presentation and MT Output Scoring

During the interview process, the selected abbreviation examples were presented consecutively to healthcare staff for their responses. A blind study approach was considered essential to enhancing interview response quality on the basis that, in addition to more closely mimicking the types of situations liable to occur in real life, taking such an approach would assist in safeguarding against priming or increasing the risk of social desirability bias among participants. The full interview questions are listed in the Appendices below.

Participants were first shown the ST on a blind basis, their being neither alerted to the presence of potentially higher-risk phenomena nor given any prior explanation as to the content beyond its pertaining to patient medical documentation and provenance from authoritative clinical data. Furthermore, the ST was framed in terms of a scenario specific to what each participant might encounter in the course of their particular professional role. This contextualisation was considered important to ensuring that participants fully grasped the aim of the exercise beyond the purely linguistic and increased their confidence in being able to provide relevant and comprehensive information based on their approach to and/or experience of such documentation in real-life settings.

Participants were first asked what information they might look to identify in the ST and for what purposes (Q1) to gain an insight into their particular priorities and immediate thought processes (e.g., identification of specialty, degree of clinical urgency, etc.), should these vary by professional role. They were then presented with the corresponding MT in English, retaining access to the ST for reference purposes, and asked what action they would take in the event of identifying a potential error in the example MT output (Q2). Where participants were unable to identify any errors in the given examples, they were permitted to answer the question on a more general basis, again with a view to understanding their thought processes and practices when managing such potential risks in real life.

Participants were then asked to score their degree of confidence in using the MT output in a real-life professional situation, along with the reason for their score (Q3). The following 6-point Likert scale was applied: 0 = ‘not at all confident’; 1 = ‘slightly confident’; 2 = ‘somewhat confident’; 3 = ‘fairly confident’; 4 = ‘confident’; 5 = ‘completely confident’.

This first round of scoring was followed by a discussion with the participant as to the general content of the MT (e.g. the specialty, scenario, and the nature of the information being conveyed), the SL meaning of the centrally problematic abbreviation in question and its corresponding translation in English, and the potential clinical impact of its remaining untranslated in the MT output on care delivery. On being given this information, participants were invited to share their thoughts on these elements. They were also at liberty to comment on any other abbreviations or higher-risk phenomena they had identified during or as a result of this discussion, given the scope this provided for gaining insight into any changes in their perceptions of the MT and/or their degree of risk awareness.

Following the discussion stage, all participants were invited to provide another confidence score and corresponding reason for their score (Q3), even if this remained unchanged. Medically qualified participants were asked to provide a score for the degree of potential risk to patient safety they perceived to be present in the MT and asked to provide a reason for their score (Q4). Only medically qualified participants were invited to respond to Q4 in order
to preserve the clinical authority of the responses. A 4-point Likert scale adapted from Brunetti et al. [5] was applied accordingly: 0 = 'no capacity to cause harm'; 1 = 'capacity to cause harm'; 2 = 'capacity to contribute to or result in temporary harm requiring intervention'; 3 = 'capacity to contribute to or result in death'.

Finally, all participants were given the opportunity to add any closing comments prior to concluding the interview. This was phrased as an unofficial open question so as to elicit as natural a response as possible from participants.

6 Interviews: Data Example Results

As interviews remain ongoing, only a small sample have been fully analysed and are included in this paper accordingly. These have nonetheless yielded interesting insights in regard to participants’ awareness of the risk present in the examples provided. It should be noted that, due to the depth of analysis applicable to each participant, it was necessary to exclude the results for the ST alone (Q1). The core findings on participants’ responses to the MT output are presented according to the order in which the interview was conducted.

Participant 1 (IV1): Referral Hub Supervisor

IV1, a native European Portuguese speaker, had worked in primarily administrative roles within the NHS, including Community Outpatients and A&E, over almost ten years. While their current and prior roles were not directly clinical in nature, they noted having ‘some clinical background as I was in the ambulances [...] but my main role was a coordinator in A&E’, describing this as ‘admin with the knowledge of clinical’. During the course of the interview, IV1 stated that they could ‘understand more or less Spanish’ but not French. Both texts were presented to IV1 in the context of a patient arriving in A&E with medical information important to their continuity of care.

Example 1: ‘AO’

Q2: IV1 placed particular emphasis on their understanding that, not being medically qualified, ‘it wouldn’t be my place to make a clinical decision’ in regard to the MT’s reliability and as such would pass the text on to a clinician. They added that they would escalate it and seek an interpreter or translator ‘to be sure that this is the right translation’.

Q3 (confidence scoring prior to discussion): Of particular note was IV1’s mistrust in the MT’s reliability, seemingly owing to their unfamiliarity with the medical terminology used (‘pinhole’), for which a register shift from expert-to-expert to expert-to-lay was instead perceived as having occurred: ‘[…] it doesn’t sound clinical [or] professional […] it sounds like the type of language you would be talking to patient for them to understand’. IV1 otherwise stated that ‘all the rest, it does seem spot on […] I’m really very, very impressed […]’. A score of four (‘confident’) was assigned, with IV1 noting that the ambiguity surrounding ‘pinhole’ was the main reason for questioning the MT’s fitness for purpose.

Discussion: IV1 indicated that the numerical figure preceding ‘AO’ had led them to believe that this referred to a medication name rather than a body part, thus not initially catching their
attention as higher-risk: ‘[...] I thought it was a saline, a sort of medication, I just jumped it ahead’. They added that, without the preceding full term from which to glean more context for ‘AO’, ‘you wouldn’t know [what it was]’ and, in attempting to identify it as a medication, clinicians ‘would be wasting [...] hours no end, because they would not find it’. This discussion appeared to galvanise IV1 into examining other abbreviations present in the MT more closely, with their independently making the observation that ‘FO’ (‘fondo de ojo’/’[eye] fundus’) likewise remained untranslated.

**Q3 (confidence scoring following discussion):** IV1 reduced their score from four (‘confident’) to three (‘fairly confident’) on the grounds that ‘in this situation [...] ['AO'] is important, but it’s not essential’ according to their perception that the clinical picture was normal, though they had interpreted the text as pertaining only to one eye, rather than both.

**Example 2: ‘DRS’**

**Q2:** The same answer as previously applies to this question.

**Q3 (confidence scoring prior to discussion):** IV1 indicated greater confidence in this MT than for Example 1, seemingly due to its more cohesive reporting structure and the perceived accuracy of the terminology and/or lexis used: ‘[...] it makes sense, what I’m reading in English’ and ‘the words that are translated seem more professional [...] it seems a proper report, to be honest with you [...] I would believe in this’. On independently focusing on ‘DRS’, they noted that they neither understood the abbreviation nor were able to decipher its meaning by reversing the order of the letters. Despite this, the MT was assigned ‘a confident four’.

**Discussion:** While examining the ST, IV1 initially misread the patient as having been referred to Psychiatry on the basis of a cardiac cause having been ruled out, rather than its having already been determined that they had no known psychiatric history. On clarifying this point during the discussion stage, IV1 speculated on the possible meaning of ‘DRS’: ‘I don’t know exactly what it stands for, but I would say something on those lines of anxiety, panic attacks, that are still showing the pain and the pressure on the chest, giving the idea of a heart attack or a stroke, for example’. However, as no such psychiatric history was present, IV1 ultimately remained unsure as to its significance.

**Q3 (confidence scoring following discussion):** IV1 indicated that they would not feel confident using the MT owing to the unresolved ‘DRS’: ‘[...] in health, there’s no assumptions, you need to have facts [...] it needs to be clear’. They further added that ‘[...] I do not trust Google Translate, because you are quite confident in most of the text and then it comes [to] the bottom and you think, ‘Oh...’ [...] ‘Can I trust the rest of what is above? Maybe not’. Notably, due to this ambiguity surrounding the patient’s ongoing symptoms, IV1 not only revoked their previous score but ‘would go even lower, because in this situation, it could be [a] more serious outcome [...] [she] still has the pain, so that needs to be investigated. And that is not clear’. Their score was reduced from four (‘confident’) to two (‘somewhat confident’) on the basis that they no longer trusted the MT: ‘[...] that took my confidence completely away [...] it makes sense what is above, but does it? [...] I don’t trust it anymore’.
Participant 2 (IV2): Advanced Clinical Practitioner

IV2, a native English speaker, had worked extensively in emergency care nursing in the NHS for nearly thirty years across a variety of specialties and care settings and had since begun practising as an Advanced Clinical Practitioner¹. In terms of language knowledge, they indicated being able to ‘make out bits and bobs that [...] you could pick up on’ in Spanish but were otherwise unable to speak or read it. Throughout the interview, they appeared to recognise French words with greater ease, though did not claim to have any formal knowledge of the language. Both texts were presented to IV2 in the context of a patient in need of treatment, accompanied by medical record information from their home country.

Example 1: ‘AO’

Q2: IV2 immediately identified ‘AO’ as unrecognised but attributed this to unfamiliarity with the specialty in question rather than a translation issue (‘I don’t know what ‘AO’ is [...] because I’m not an Ophthalmology specialist’), noting that they would ‘refer to a specialist to be on the safe side’.

Q3 (confidence scoring prior to discussion): IV2 assigned the MT a score of two (‘somewhat confident’) on the basis that it enabled them to establish a rough clinical picture, though this was hampered by their own unfamiliarity with the specialty in question. Interestingly from a comparative confidence perspective and in view of the subsequent cardiological French ‘DRS’ example, they also noted that ‘If it was something like a cardiac presentation, for example, I’d know an awful lot more, and I could be a lot more confident in interpreting the text based on what I’d seen with the patient’.

Discussion: IV2 indicated that their unfamiliarity with the specialty drew their attention to ‘AO’ and informed their decision to seek help with the MT, though they still perceived this ambiguity as being clinical in nature, rather than a translation issue: ‘[...] I thought ‘I don’t know what that means’, but Ophthalmology would look at that and [...] probably make an educated guess [...]’. Furthermore, on discussing the lack of preceding full term with which to decipher ‘AO’ by comparison with other abbreviations in the text, IV2 highlighted the fact that ‘If there’s no context to it [...] you just don’t know what it is’.

Q3 (confidence scoring following discussion): IV2 indicated that they would retain their confidence score of two (‘somewhat confident’), seemingly on the basis of most of the MT enabling them to establish a broad clinical picture. However, they emphasised having taken note of the ambiguity arising from the lack of context preceding ‘AO’ in their assessment: ‘[la presión intraocular] you can, sort of, interpret that, and see [...] that relates to the words before, but just the abbreviation, that has no context whatsoever, does it?’.

¹ Health Education England (HEE) defines Advanced Clinical Practice as being ‘delivered by experienced, registered health and care practitioners. It is a level of practice characterised by a high degree of autonomy and complex decision making. This is underpinned by a master’s level award or equivalent that encompasses the four pillars of clinical practice, leadership and management, education and research, with demonstration of core capabilities and area specific clinical competence’. [10]
Q4 (potential risk to patient safety scoring following discussion): The MT was assigned a score of two (‘capacity to contribute to or result in temporary harm requiring intervention’) based on the need to seek assistance with the translation to avoid negatively impacting on patient safety, adding that ‘you would definitely need to get help with this’.

Example 2: ‘DRS’

Q2: IV2 immediately identified ‘DRS’ as unrecognised, along with the inconsistent translation of gender throughout the MT (alternating between male and female) and cardiology-specific terminology with which they were unfamiliar (‘resténose’/’restenosis’). Again, they indicated that the MT may be useful for obtaining valuable information but would be mindful of the errors identified in assessing its reliability for clinical purposes: ‘[...] it does give me important information, but there are certain things in there that I would think, ‘That’s not right’’. They emphasised the importance of integrating the information provided by the MT into a wider clinical assessment of the patient and any attendant investigations, approaching the text with a degree of caution accordingly: ‘[...] I would take some of the information from this, but I would have definitely got it interpreted and maybe referred to Cardiology if I was concerned about the patient, but [...] I wouldn’t rely on it too heavily, at all’.

Q3 (confidence scoring prior to discussion): Despite the untranslated ‘DRS’, a confidence score of two (‘somewhat confident’) was assigned to the MT on the basis of its enabling IV2 to establish an overall picture of the patient’s medical history, though its use in practice would fit within the wider framework of their clinical assessment of the patient.

Discussion: IV2 indicated that despite obtaining a general overview of the clinical context from the MT, the non-translation of ‘DRS’ remained centrally problematic to treating the patient’s symptoms and would therefore need to be resolved due to its potential criticality in clinical terms: ‘I would want help with that, because I don’t know what it is, and it might be very important [...] [‘DRS’ has] been investigated, nothing found, but what is it?’.

Q3 (confidence scoring following discussion): IV2 again expressed an appreciation for the clinical background provided by the MT, though within the context of its being complementary to their wider clinical assessment: ‘[...] you would have confidence, in a way, that you’re dealing with a certain situation’. Despite this, their score was reduced from two (‘somewhat confident’) to one (‘slightly confident’) in acknowledgement of the need to resolve ‘DRS’, for which they indicated referring to a French-speaking colleague or Language Line [an interpreting and translation service used across the NHS].

Q4 (potential risk to patient safety scoring following discussion): IV2 assigned the MT a score of three (‘capacity to contribute to or result in death’) owing to the potentially significant nature of the medical issue at hand and the ambiguity surrounding the centrality of ‘DRS’ to determining the ongoing symptoms and prioritising patient safety; ‘Because it’s a cardiac presentation [...] [it’s] probably more critical to understand the interpretation because it could be something that you could get very wrong’.
In closing, IV2 indicated that the interview had given them pause in considering the use of abbreviations in their own professional practice, particularly given the ubiquity of such features in both verbal and written healthcare contexts. However, time was noted as a contributor to their continued use: ‘We use abbreviations all the time at work [...] and if I was to document, should I be using abbreviations or should I be using the full text? And I think I should be using the full text. But time is a factor [...] [Abbreviation use] is a language in itself’.

7 Discussion of Preliminary Findings

7.1 Impact of MT Use with Patient Medical Record Documentation

The examples analysed in Section 4 indicate that certain medical abbreviations, particularly those for which no surrounding context in the form of a preceding full term is provided in the ST, run the risk of remaining entirely untranslated in the corresponding MT output. In the case of ‘AO’, a correct translation should have reflected the directly equivalent abbreviation available in either Latin or English. By contrast, for ‘DRS’, the failure of the MT to provide an appropriate translation may be attributed to the fact that no such directly equivalent abbreviation exists in English, in addition to the absence of preceding context in the ST. Given the higher-stakes setting and the centrality of both abbreviations to comprehension of their respective clinical contexts, their non-translation not only implies an error at the linguistic level but also a potential increase in risk to patient safety.

7.2 Awareness among Healthcare Staff of Risks Posed by MT Use with Patient Medical Record Documentation

While both participants initially underestimated the relevance of the selected abbreviations to the MT’s clinical fitness for purpose, their respective responses indicated an interesting difference in reasoning as to the influences on their perceptions in this regard. IV1’s view of MT reliability appeared to be more heavily influenced by the perceived text quality at a broader linguistic level and its conformance to typical reporting structures, as well as the cohesiveness of each text as a whole. Further to this, they acknowledged having originally overlooked the untranslated ‘AO’ owing to the assumption that it represented a different type of information. IV2’s view of this reliability appeared to depend more heavily on their familiarity with the specific specialty, rather than an awareness of a translation issue present in the MT. This was reflected in the fact that, while IV2 immediately identified both abbreviations as unrecognised, their initial assumption with regard to ‘AO’ was that this was due to their unfamiliarity with ophthalmology rather than its remaining untranslated in the MT output. Following discussion, both participants appeared to adopt a more critical eye from a translation perspective, with IV1 in particular paying closer attention to other untranslated abbreviations in Example 1 and both IV1 and IV2 focusing more closely on the impact of the lack of a preceding full term on the (non-)translation of both ‘AO’ and ‘DRS’.

In terms of changes to participants’ degree of risk awareness following discussion, IV1 initially expressed less concern over the non-translation of ‘AO’. This was seemingly based on their perception of the patient’s overall clinical stability. However, following discussion of ‘DRS’, they indicated a significant drop in trust in the MT output owing to the potentially
serious clinical implications arising from its non-translation, this distrust extending to the reliability of the entire MT as a result. By comparison, IV2 appeared to adopt a more integrative approach to assessing risk in the MT output in terms of its allowing them to establish an overall picture of the patient’s background within the framework of their wider clinical assessment. However, at no point were the errors and/or ambiguities present in the MT dismissed by IV2 as a result of this approach; to the contrary, it was emphasised throughout that these would need to be identified and addressed in order to preserve patient safety, particularly given the ambiguous nature of the ongoing symptoms indicated by ‘DRS’.

In regard to management of the potential risks posed by using the MT in a professional setting, both participants demonstrated sound awareness of the need to confirm the content of the MT to ensure clarity of comprehension and preserve patient safety, whether through deferral to a clinical specialist and/or professional language services. In particular, IV2 emphasised the fact that while they considered the MT to contain potentially important information that should not be disregarded outright based on the ambiguities and/or errors present, the MT output would require both double-checking and integration into their own clinical assessment of the patient, rather than relying on it in isolation. Furthermore, IV2 commented that the interview had inspired them to reconsider their own practices in terms of abbreviation usage in a real-life setting, despite the time constraints present in their work environment.

8 Conclusions and Future Work

Both participants demonstrated an encouraging degree of risk awareness in their respective approaches to managing ambiguities and/or errors present in patient medical record information translated with MT. Nonetheless, it would appear that there remains a need to raise awareness among healthcare staff concerning the closer consideration of unrecognised medical abbreviations in MT output, particularly where these are inadequately translated or not translated at all, and may increase the potential for risk to patient safety as a result. This is especially relevant to medical abbreviations lacking in preceding context in the ST, the non-translation of which may impact negatively on care delivery due to their being overlooked by healthcare staff as less immediately relevant to clinical assessment than more readily recognisable medical terminology, misinterpreted as other types of information, or simply assumed to belong to another specialty. The difference in reasoning as to the influences on each participant’s perceptions of the MT’s fitness for purpose also raises interesting questions as to the degree to which future participants may interpret text quality at the broader linguistic level (that is, in terms of lexis and/or terminology, conformance to expected genre conventions, etc.), along with the extent to which this perception influences their awareness of the risks presented by both medical abbreviations and their (non-) translation with MT. These findings are currently being followed up by interviews with a further 18 participants and shall be subject to more comprehensive analyses accordingly.
9 References

10 Appendices

Interview questions

Q1 (ST only): What kind of information in the documentation, if any, would you be looking to identify, and for what purposes?

Q2 (MT with ST available for reference): If you identified, or thought you may have identified, an error in the machine translation output, what would you do?

Q3 (confidence scoring requested from all participants prior to and following discussion): How confident would you feel about using this machine translation output in a real-life professional situation? Please give a score for your degree of confidence and explain your answer.

Q4 (potential risk to patient safety scoring requested from medically qualified participants only, following discussion): Please provide a score for the degree of potential risk to patient safety you would consider this machine translation output to pose if used in a real-life professional situation. Please explain your answer.