



Communication Crossroads

Adoption of electronic data capture in emerging countries has initially been slow. Karen I Politis Virk at Language Connections explains how increased exposure to the internet, technological advances and unified translation guidelines can help to ease the transition

With the expansion of clinical research into emerging countries, many companies are pushing for global electronic data capture (EDC) adoption in an effort to improve the efficiency and accuracy of patient data collection methods. Although there are many advantages to converting to a paperless system, there are several new challenges that have arisen in the process. Limited IT infrastructure, inadequate training in EDC technology and language barriers have all been contributing factors to the relatively slow progress of EDC adoption in clinical trials globally.

In order to overcome the issue of language, multilingual materials must be developed. Translating eClinical materials for the collection of patient data, however, requires extensive expertise. For one thing, the accuracy of patient data is directly related to translation. Developing a multilingual EDC system has been hindered by the vast linguistic and cultural differences among patient populations in international clinical trials, as well as significant structural differences between English and many of the languages into which materials must be translated. In addition, patients with limited reading proficiency pose further challenges that must be addressed.

Although EDC is most prevalent in US and western European clinical trials, increased exposure to the internet and more recent EDC technology such as wireless devices has facilitated its adoption in emerging countries with limited infrastructure. In addition, the use of a universal system for computerised scripts has helped expand EDC globally. If there is a concerted effort to apply unifying translation guidelines, language barriers can also be overcome. Current industry projections estimate that EDC is now used in approximately 50 per cent of all clinical trials (1).

GLOBALISATION CHALLENGES

Several challenges have resulted from the growing prevalence of global site users and increased outsourcing

of clinical trials. For example, while the global expansion of trials presents new opportunities, it has also increased the demands placed on EDC and IT infrastructure. Until recently, many emerging countries did not have the IT infrastructure to support conversion. In some cases, computers were not available in clinical settings and training personnel in EDC systems was not yet well established. According to 2009 statistics, between 10 and 14 per cent of trial sites in Russia, China and Latin America required computers and/or training compared to one to two per cent of sites in the US (2).

In some countries, the cost of establishing EDC technology is one of the biggest barriers to conversion. For example, although the majority of western companies conducting trials in China use EDC, most Chinese companies do not. One of the main reasons for this is the cost of adopting EDC technology versus the low cost of Chinese labour (3). Although unreliable internet access and inconsistent electrical output have been a common problem, even more so in some parts of Africa, mobile phones and wireless handheld devices are helping to overcome this issue. As a result, despite the fact that North America and Europe remain ahead, new markets are gaining pace. However, conversion has been slow due to a number of factors. In Japan infrastructure is not an issue, but many of the same factors which inhibited EDC adoption in the past – such as converting to Japanese script and the complexity of training physicians – continue to contribute to the slow pace of adoption. As of May 2008, 40 per cent of clinical sites in Japan were using EDC for some or all of their trials (4).

Another challenge in adopting EDC has been converting various other written scripts. Until recently, this made EDC adoption in Japan difficult since Japanese uses three different character sets, Kanji, Katakana and Hiragana, each of which has a different width and size. In India, the situation is yet more complicated as different languages use different scripts. For example, Hindi, Marathi and

“ In some countries, the cost of establishing EDC technology is one of the biggest barriers to conversion. For example, although the majority of western companies conducting trials in China use EDC, most Chinese companies do not ”

Nepali share a common script (Devanagari), while Telugu, Kannada and Tamil have their own scripts. Chinese languages are written using two different character sets, Traditional (common in Taiwan and Hong Kong) and Simplified (primarily used in mainland China). With the development of the Unicode System – a universal character encoding system that is used primarily to represent various characters and symbols expressed by different writing scripts – this issue has largely been resolved. Many scripts, such as Korean, Japanese and Chinese have been adapted to the Unicode System.

In order to successfully expand EDC internationally, linguistic differences must also be addressed. This means that companies must provide technical support 24 hours a day, seven days a week in any given language, and materials such as user manuals, web-based interfaces, and automated materials for the collection of patient data (using case report forms (CRFs), patient reported outcomes (PROs), patient questionnaires and patient diaries) must all be translated into required native languages. It should be noted that although inconsistent standards across clinical sites and inadequate training of personnel in EDC use are considered among the most critical factors affecting data quality in emerging countries, the vast linguistic and cultural differences among patient populations in global trials also have a significant impact.

TRANSLATION CHALLENGES

Although conversion from paper-based methods to EDC technology has significantly improved the level of efficiency and accuracy of clinical data collection, the role that language and culture play must not be underestimated. The success of multilingual EDC systems relies heavily on accurate translation. In addition to the vast linguistic differences among patient populations, many of the languages into which materials must be translated differ significantly from English – the language in which the majority of EDC systems have been constructed. Often, differences between grammar and sentence structure can be significant, making translation challenging even for experts. Moreover, due to the critical nature of the material it is essential that translation is performed using established methods that typically involve several steps.

In order to translate materials for eClinical patient data, the text must be first separated from the code, translated, reviewed, then reinserted and checked again to ensure that the codes are properly aligned. Translation commonly is conducted by an experienced team of native speakers who perform forward and back translations, and a separate

team that then reviews the translation. Following translation and localisation of the text, an on-site linguist must check the translated software version. Often as a final step, the translation must undergo pretesting by individuals from the target population. Although this method is tedious and often quite time consuming, it is by far the most effective. However, there are several factors associated with obtaining accurate patient data from international sites that must be addressed. These include a lack of unifying translation standards, significant structural differences between languages, vast linguistic and cultural differences among patients, and variable patient reading proficiency.

Lack of Unifying Translation Standards

Although translation standards exist, they are not always fixed. For example, there are flexible standards as to what is acceptable across different local languages in some countries. A clinical study conducted on psychiatric patients in India which required the translation of automated psychiatric assessment patient questionnaires into several Indian languages (including Hindi, Punjabi, Marathi, Malayalam, Telugu, Kannada and Gujarati) demonstrates this issue. According to one of our own senior project managers responsible only for reviewing previously translated versions, there were “flexible or fluctuating standards as to what was considered correct grammar and spelling across these languages”. In addition to variation within the same country, sometimes there are variable standards between different countries or regions. Moreover, although most countries have established guidelines, they may not be completely uniform when applied to languages that are less commonly translated. This is due to the fact that it may be challenging to locate experts in those languages.

Linguistic Variation

Another issue is related to the range of linguistic variation, or numerous regional dialects, that must be accounted for in some countries. For example, although Castilian Spanish is the only form of written Spanish, each country and region in Latin America has its own variation of spoken Spanish and cultural elements that influence the way the language is used. Mexican Spanish differs significantly from Argentine Spanish, and both in turn differ from Castilian Spanish. A look at regional dialects spoken throughout Mexico alone demonstrates the complexity of this issue – currently, there are more than 50 native Mexican languages considered to be Mexican Spanish dialects (5). Spanish-speaking populations therefore in central Mexico, the Yucatan Peninsula, or northern Mexico all speak a different variety of Spanish as a result of different indigenous influences (6). Many borrowed words from indigenous languages in each region have

“ Although conversion from paper-based methods to EDC technology has significantly improved the level of efficiency and accuracy of clinical data collection, the role that language and culture play must not be underestimated ”

replaced Spanish words, such as names for food or everyday items. Thus, expert translators who are native speakers that are familiar with these differences are best equipped to address them.

Identifying Proper Language Equivalents

In addition, many languages into which materials must be translated are quite unrelated to English – the language in which most EDC materials, such as automated patient questionnaires, have been constructed. As a result, identifying language equivalents can be particularly challenging. A look at linguistic differences between Chinese and English, for example, makes this evident. Chinese uses characters rather than letters and is composed of two separate character sets. Unlike English, a single character typically represents multiple grammatical forms (for example *treat*, *treating*, and *treatment*). Thus, a sentence constructed in Chinese may be considerably different from the English one. Another example is that abbreviations do not exist in Chinese, and thus each word must be translated first into Chinese and then abbreviated.

The Russian language presents its own challenges, especially when translating from English into Russian. Although the Cyrillic alphabet is not entirely unrelated to the Latin alphabet, Russian differs from English in several aspects including grammar and word order, and has one of the most complicated punctuation systems (7). Another language which is structured very differently from English is Turkish. The Turkish language is composed of significantly less words (90,000) than English (350,000). This is a good indication as to why Turkish translations are often significantly shorter (8). Often translators must come up with the best closest equivalent since there may not be a direct language equivalent in some cases.

Several multinational studies have demonstrated how translating eClinical materials used in the collection of patient data can be problematic. In one study aimed at adapting PROs from English into 14 different languages, numerous linguistic issues became apparent during translation. For example, in translating a particular item, “Under each heading, please tap the ONE box that best describes your health TODAY. Please tap on the scale to indicate how your health is TODAY”, translators encountered several issues. A Hungarian interviewer reported that the word ‘press’ had to be substituted for ‘tap’ since patients were not able to provide answers by ‘tapping’. The eastern European translators reported that ‘section’ or ‘groups’ had to be substituted for ‘heading’ since there was no correct literal translation. In Latin American Spanish, the emphasis of ‘tap ONE box’, meaning only one, became lost in translation due to the requirement for an article in Spanish (*una* or *la*) (9). Despite these differences, when established translation methods are applied by experts in the field, these types of problems can be overcome.

Cultural Biases and Every Day Practices

Culture has a significant impact on the behaviour and perception of individuals not only in each country, but also within a specific community. Thus, translating materials

for patient-generated or reported data requires a clear understanding of any cultural aspects that can interfere with patient comprehension. For example, how a patient perceives and deals with pain depends on the culture that they are a part of. Data from several international cancer studies indicates that pain is typically underreported due to local cultural attitudes. Therefore, in order to standardise clinical data such as pain rating and ensure accuracy, cultural attitudes towards pain must be well understood. In addition, in countries where natural therapies are commonly used in conjunction with or in place of conventional medicines, quality of life or health assessment questionnaires must include a question dealing directly with the possibility of natural treatments used by patients during the course of a study.

Furthermore, what is meaningful in one language can be completely void of meaning in another as a result of cultural differences. Every day items or daily activities may differ vastly from culture to culture, as well as associated vocabulary. For example, one study focused on diagnosing clinical depression in geriatric patients in Malaysia reported that one of the questions in the study questionnaire had absolutely no discriminatory value in differentiating cases of depression among Malay patients: “Do you prefer to stay at home, rather than going out and doing new things?” Since in Malaysia the main emphasis is on family living rather than individual lifestyle, patients had a higher tendency to give positive response to this item, despite their depression (10). For this reason, translation must incorporate cultural context to account for cultural differences.

Limited Reading Proficiency

In some countries where there are multiple official languages or several dominant spoken languages, it is important to establish the patient’s reading proficiency, as well as their native language. Even in countries where there is only one official language, EDC adoption may be affected by the fact that many patients have limited reading comprehension.

Many ethnic minorities and indigenous groups, for example, tend to have limited proficiency in the official language of their country. In addition, certain patient populations may have a higher rate of illiteracy. Thus, special measures must be put in place to ensure adequate patient understanding. In addition, many patients with lower economic status have not had the opportunity to become computer literate and may need additional assistance in using EDC technology.

National literacy rates in many emerging countries have improved. However, statistics vary greatly depending on the various demographics of a population. For example, women tend to have a higher rate of illiteracy, as do ethnic minority populations and individuals from rural areas. Furthermore, although some patients may not be considered illiterate, they may have a limited understanding of medical concepts or inadequate reading comprehension.

In many Latin American countries, women and indigenous populations have reportedly lower reading comprehension

and higher illiteracy rates than the national averages (11). Brazil's illiteracy level is higher than most other Latin American countries, and although only around 10 per cent of the total Brazilian population is officially considered illiterate, other statistics indicate over 30 per cent of the population may not have the capacity to read and write texts (12). Moreover, there are large regional and ethnic differences.

According to the *CIA World Factbook* March 2007 statistics, India's adult literacy rate is estimated at 61 per cent. However, there is a great amount of variability throughout different regions and among different socio-economic groups. In addition, illiteracy is significantly higher among women (13). Although the reported literacy rate among adults in South Africa is 86 per cent, this rate varies significantly among different ethnic populations according to *CIA World Factbook* 2005 statistics.

Lack of healthcare education, poor reading comprehension and limited computer skills make conversion to EDC among these populations more challenging. Beyond the issue of translation, however, additional measures must be taken so that ethical standards are maintained.

CONCLUSION

The use of EDC systems in international clinical studies has greatly facilitated the management of clinical data by improving data accuracy, and increasing the standardisation and efficiency of data collection. With the expansion of clinical trials into emerging regions, there is a growing demand to convert to EDC technology. However, there are several factors which have slowed the process of adoption, including limited IT infrastructure, inadequate training in EDC use, and language barriers. To date, the majority of EDC systems are English-based. The challenge therefore lies in translating them into the large number of languages native to patient populations in emerging countries, and overcoming the fact that differences between some of these languages and English may be significant.

Several factors can impede the success of developing multilingual eClinical materials if they are not addressed. These include a lack of unifying translation standards, structural differences between languages, vast linguistic and cultural patient differences, and variable patient reading proficiency. Translations must be linguistically validated and culturally adapted for each patient population. Moreover, patients with limited reading proficiency must be given additional attention.

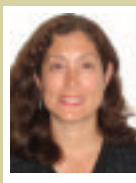
The success of EDC technology in clinical trials lies in proper implementation. If multilingual eClinical materials are to be effective, they must undergo a standardised translation process. This process, although timely and tedious, has been shown to be the most effective. When proper methods are applied, the language barrier can be overcome. However, familiarity with the specific factors that affect the transmission of correct information,

patient comprehension, and the accuracy of patient data is also essential.

References

1. CenterWatch, EDC Adoption in Clinical Trials: A 2008 Analysis, www.bio-itworld.com/EDC-ADOPTION.aspx
2. Schatzma BD, eClinical solutions in China, *International Clinical Trials* 15: pp50-52, 2010
3. Albani CR and Tabata Y, Japan – Thinking global, acting local, *Good Clinical Practice Journal*, 15(5): pp13, 2008
4. Hanaman D and Vollkommer R, Optimising and supporting remote data capture, *Health Sciences Visions*, 2009, www.healthsciencesvisions.com/articles/optimising-and-supporting-remote-data-capture?page=0,3
5. Schmal J, The linguistic diversity of Mexico, *Houston Institute for Culture Educational Literature*, 2004, www.houstonculture.org/mexico/ling.html
6. Latin American Spanish: Mexican Dialects, *Omniglot*, www.omniglot.com/language/articles/latin_american_spanish.htm
7. English to Russian Translation; Some Facts, www.language-translation-help.com/english-to-russian-translation.html, Accessed online April
8. Ozolins U, Back translation as a means of giving translators a voice, *The International Journal for Translation & Interpreting Research*, 1(2): pp1-13, 2009, <http://transint.org/index.php/transint/article/viewFile/38/55>
9. Ashcroft-Jones A, Furtado T and Wild DJ, The translation and linguistic validation of the EQ-5D electronic version, *Oxford Outcomes*, 2009, www.oxfordoutcomes.com/library/conference_material/posters/PRO/PRO%20The%20Translation%20and%20Linguistic%20Validation%20of%20the%20EQ-5D.pdf
10. Teh EE and Hasanah CI, Validation of Malay version of geriatric depression scale among elderly inpatients, *Priory*, 2004 <http://priory.com/psych/MalayGDS.htm>
11. Latin America and the Caribbean: Gender illiteracy, patterns of inequalities in adult education, *Latin American Ministries of Education*, 2003, www.crefal.edu.mx/biblioteca_digital/temas_especializados_epja/alfabetizacion/illiteracy_adults_gender_latinamerica.pdf
12. Brazil – Poverty illiteracy and income concentration, *Encyclopedia II*, www.experiencefestival.com/a/Brazil_-_Poverty_illiteracy_and_income_concentration/id/4868958
13. Kohli A and Basu A, India, In Kesselman M *et al* (eds), *Introduction to Comparative Politics*, p283, 2009

About the author



Karen I Politis Virk has been Director of Biotech and Pharmaceutical Research at Language Connections in Boston, MA, since December 2007. Prior to her position at Language Connections, Karen worked in research and development at several pharmaceutical and biotech companies as a research associate in molecular biology, including MicroGeneSys, Wyeth Ayerst, TransKaryotic Therapies, and Cubist Pharmaceuticals. Her experience includes the development of products for human clinical trials, particularly the human AIDS vaccine.

Email: karen@languageconnections.com