

# COUNTING ON DATA VALIDITY IN MULTINATIONAL CLINICAL TRIALS? NUMERAL SYSTEMS IN PATIENT-REPORTED OUTCOME ASSESSMENTS: GUIDANCE FOR LINGUISTIC VALIDATION

Tim Poepel, PhD, Rachael Browning, BA Hons., Rebecca Israel, MS, Oliver Delgaram-Nejad, PhD, Payton Ramsey, MPH, Allyson Nolde, MPP, Chryso Hadjidemetriou, PhD, Elizabeth McCullough, MA



## INTRODUCTION:

Western Arabic numerals (i.e., 1, 2, 3) are the most commonly used symbols for numbers, and almost exclusively appear in items or response options in patient-reported outcome assessments (PROs). Linguistic validation (LV) is the process of translating and culturally adapting PROs resulting in comprehensible, patient-friendly translations adapted for use in specific countries, cultures, or populations. Despite wide use and recognition of Western Arabic numerals, some languages have and preferentially use idiosyncratic numbering systems in certain contexts. This work addresses a gap of information regarding such languages, their numeral systems, and the availability of guidance and feedback regarding how and when different numeral systems are used, with the broader goal of improving LV of numeric text and promoting awareness of the necessity of accommodating variation in the use of numeral systems.

## METHODS:

We surveyed 124 linguists representing 63 languages (see Table 1). Questions asked about the existence of unique numeral systems, contexts for use of Western Arabic and unique systems, preference for system use in daily life, and preference in writing vs typing, plus open-ended feedback on numeral systems in general.

Table 1: 63 Languages in the sample

Afrikaans	Hebrew	Setswana
Arabic	Hiligaynon	Slovak
Bosnian	Hindi	Slovene
Bulgarian	Hungarian	Spanish
Catalan	Igbo	Swahili
Chinese - Traditional	Ilocano	Swedish
Chinese-Simplified	Italian	Tagalog
Croatian	Japanese	Tagalog
Czech	Kannada	Tamil
Danish	Kanuri	Thai
Dutch	Korean	Tswana
English	Latvian	Turkish
Filipino	Malay	Ukrainian
Finnish	Malayalam	Urdu
Flemish	Polish	Venda
French	Portuguese	Vietnamese
Georgian	Romanian	Xhosa
German	Russian	Xhosa
Germany	Sepedi	Xitsonga
Greek	Serbian	Yoruba
Hausa	Sesotho	Zulu

Table 2: Other Numeral Systems Identified in the Sample

Language	Numeral System
Arabic	٠ (٠), ١ (١), ٢ (٢), ٣ (٣), ٤ (٤), ٥ (٥), ٦ (٦), ٧ (٧), ٨ (٨), ٩ (٩), ١٠ (١٠)
Chinese Simplified and Traditional - Financial	零, 壹, 貳 or 貮, 參 or 叁, 肆, 伍, 陸 or 陆, 柒, 捌, 玖, 拾
Chinese Simplified and Traditional - Ordinary	一, 二, 三, 四, 五, 六, 七, 八, 九, 十
Georgian	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Greek	Α, Β, Γ, Δ, Ε, Ζ, Η, Θ, Ι
Gujarati	૦, ૧, ૨, ૩, ૪, ૫, ૬, ૭, ૮, ૯
Hebrew	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Hindi	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Japanese	一, 二, 三, 四, 五, 六, 七, 八, 九, 十
Kannada	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Korean	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Malayalam	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Tamil	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Telugu	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Thai	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Urdu	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Vietnamese	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

## RESULTS:

17/63 (27%) of languages in our sample reported some use of a system other than Western Arabic numerals (see Table 2). An additional 8 languages reported use of Roman numerals in some contexts, bringing the number of languages using two systems in the sample to 25 (40% - see Table 3).

Table 3: Languages also using Roman Numerals

Catalan	Igbo	Russian	Spanish
Danish	Italian	Slovak	Swedish

For languages using two systems, 18/25 (72%) noted contextual factors dictating use, such as legal documents (e.g., Traditional Chinese), government documents (Thai), urban vs rural target demographics (Hindi), stylistic or specialized fields (Slovak; Catalan), or formality of intended language (Chinese - Simplified, Traditional). See Table 4 for a more detailed overview of feedback on context of use.

Table 4: Detailed linguist feedback on use of numeral systems

Language	Linguist Feedback
1 Catalan	Roman numbers for kings and queens or some specialized fields
2 Chinese - Tradition and Simplified	If the document is formal, i.e. a legal document or medical report, Western numbers are normally being used. If document is for a specialized or traditional context, such as finance, translating a story or a description of a language specific issue, then another system will be used.
3 Danish	We also have Roman numbers, e.g. MMXXIV for 2024, but we rarely use them.
4 Hausa	Sometimes in written documents, the figure is written in words. Apart from this, figures are generally written using the Western Arabic Numbering system.
5 Hindi	Most of the official communications, publications, media channels etc. now use the Western Arabic numbers but some publications prefer to use the Devnagari system. Many people in rural areas also use this when writing.
6 Igbo	The Roman numerals are also acceptable in Igbo.
7 Italian	In Italian we generally use the regular Western Arabic numbers. We also use Roman numerals, though. These are commonly used in formal writing, for dates, in outlines, and in the naming of monarchs and popes. For instance: Dates: "il XIV secolo" (the 14th century) Outlines: "Capitolo II" (Chapter 2) Naming Monarchs and Popes: "Papa Giovanni XXIII" (Pope John XXIII) The use of Roman numerals adds a sense of formality and tradition.
8 Japanese	Often Kanji numbering characters are used when other Kanji is in the text, and generally used less than other number characters
9 Kannada	In Kannada there is separate numbering system. But academically and in our day-to-day communications we use Western Arabic numbering system.
10 Malayalam	Using Western Arabic predominantly
11 Romanian	Romanian only uses Western Arabic numbering system. The only difference is that for the English UK/US 1,000.5 (one thousand point five) in Romanian we use dot/ space for thousands, and comma for decimals, i.e. 1.000,5 or 1 000,5
12 Russian	Sometimes Roman numbering system is used in numbering chapters in the books or when centuries are mentioned.
13 Simplified Chinese	1. Arabic numerals: Numbers used for measurement; Numbers used for numbering, such as telephone numbers, postal codes, bus numbers 2. Chinese character numerals: Lunar calendar; Fixed words containing Chinese characters and numbers, such as the May Fourth Movement
14 Slovak	Latin numbers can be used, however, it's rare and used mostly as formatting / stylistic contexts.
15 Spanish	There are a few things for which Roman numerals are always used (centuries, Olympics, regular conferences or meetings, some anniversaries, etc.)
16 Swedish	The names of smaller integers (usually 1-12) are written with letters and inflected for grammatical gender. In Swedish, Roman numerals are also frequently used, nowadays usually when an alternative to Arabic numbers is needed.
17 Thai	Nowadays we tend to use Western Arabic numbers more. However, in some strict government documents, they still use Thai number system.
18 Urdu	Mostly use Western Arabic numerals presently
Finnish	Measurements only in metric measures, always 24 hour time.
Dutch	Dutch people get terribly confused when they have to answer questions using AM and PM - we always use 24h notation.

For both writing and typing in everyday use, 5/17 (29%) of linguists in multi-numeral system languages reported no system preference, while 3/17 (18%) preferred using the unique system, suggesting that Western Arabic numerals may be preferentially used over other systems.

All 63 languages reported familiarity with and use of Western Arabic numerals in daily contexts, and only Chinese (both Simplified and Traditional) linguists suggested a strong inclination toward using a different numeral system based on context.

## CONCLUSIONS

Results show Western Arabic numerals as the dominant numeral system, being highly preferred and used over idiosyncratic systems, but reveal a non-trivial subset of languages with unique and contextually relevant numeral systems whose use should be considered in translation and especially LV. Further work is required and planned to garner a deeper understanding of the considerations needed for this subset of languages, as well as investigation into how different numbering systems have been used to date in PROs, with the goal of producing guidelines on the use and LV of numeral system in PROs. For any PROs, appropriate numeric localization to maximize naturalness and appropriateness of the translation, and patient comprehension, is essential for collecting valid data and maintaining patient centricity.