

This is a “preproof” accepted article for *Journal of Clinical and Translational Science*.

This version may be subject to change during the production process.

10.1017/cts.2025.10078

## **Going Beyond the Technology: Considerations in Translating Electronic Case Report Forms**

Amy E. Krefman, MS<sup>1</sup>, Luke V. Rasmussen, MS<sup>1</sup>, Crystal Santillanes, MS CCRC<sup>1</sup>, Mary Beth Tull, MBA MS CCRP<sup>1</sup>, Patricia Bustamante, MPH<sup>1</sup>, Alan Kuang, MS<sup>1</sup>, Hongyan Ning<sup>1</sup>, Yaojie Wang, MS<sup>1</sup>, Barbara M. Gonzalez, BS<sup>2</sup>, Eugenia I. Manrique, BS<sup>2</sup>, Lauren B. Bonner, PhD<sup>1</sup>, Jody D. Ciolino, PhD<sup>1</sup>

<sup>1</sup>Northwestern University Feinberg School of Medicine, Chicago, IL, USA

<sup>2</sup>Miller School of Medicine, University of Miami, Miami, FL, USA

**Corresponding Author:** Amy Krefman, Northwestern University, Feinberg School of Medicine, 680 North Lake Shore Drive, Suite 1400, Chicago, Illinois 60611, Email: amy.krefman@northwestern.edu

**Keywords:** Translating; Diversity, Equity, Inclusion; Data Collection; Workflow

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

**Abstract:**

Through a series of example research studies, we illustrate processes in translating case report forms to increase language diversity in study populations while simultaneously highlighting implications for data collection and analyses. The Northwestern University Data Analysis and Coordinating Center (NUDACC) manages the translation of participant-facing study documents into languages other than English through a process that has been refined over several years, adjusting for changes in technical capabilities in electronic case report forms. This approach to manage, examine for context, and implement certified case report form translations offers an efficient workflow to streamline data capture in multiple languages.

**INTRODUCTION**

When conducting clinical research, it is important to enroll participants belonging to under-represented populations to ensure a representative sample and increase generalizability.[1] The translation of all participant-facing materials, including case report forms (CRFs), is critical to mitigate recruitment barriers for participants who may not speak English and thus retain a representative research study population. Depending on review board requirements, a certified translation service may be used [2] to translate CRFs, but ensuring usability and flexibility in function for the study participants requires an investment of time, money, and thoughtful consideration.

The Food and Drug Administration (FDA) does not have nuanced guidance on requirements for translating study materials such as informed consent documents.[2] The FDA requires more than translation alone when a patient-reported outcome serves as an endpoint; they require validation and cultural adaptation.[3] However, many studies not regulated by the FDA enroll participants across multiple languages. Despite budgeting for translations, some research teams are not able to validate or evaluate questionnaires through a rigorous cultural adaptation process. Participants tend to prefer questionnaires that undergo a cultural adaptation process over those that undergo forward-back translation (i.e., one translator translates the document into the new language, and a second translator translates it back to the original language, and the two versions are compared).[4] Cultural adaptation processes are especially helpful for identifying intangible cultural heritage terms, which are “unique expressions of cultural knowledge and practices that

are deeply rooted in a particular region or community.”[5] However, cultural adaptation is expensive, and one recognized barrier to enrolling patients from minority populations is lack of budget for translating study materials, especially for investigator-initiated studies.[6]

Additionally, different translation approaches affect the “readability, comprehension, and user preferences” of participant-facing materials.[7] These translation approaches include literal, functionalist, and equivalence-based approaches, with a preference for the functionalist approach—where a translator makes decisions based on instructions meant to better suit the needs of the target audience.[7] However, investigator-initiated studies in academia often lack personnel and resources to vet and compare translation companies on these approaches. Further, translation companies may not have the necessary contextual understanding for a clinical trial or observational study to implement a functionalist approach, and some vendors’ policies hold the requestor accountable for the translation context.

Finally, implementation of the translated study materials can be difficult. Some electronic data capture (EDC) systems can accommodate the same case report forms in multiple languages, but many do not. It is important to consider the end user experience when planning a study’s translations, which in our experience requires careful attention to interface details and functionality (e.g., the systems navigation buttons such as “next” or “submit” may also require translation).

## **Objective**

Leveraging REDCap[8,9] and a highly collaborative multi-disciplinary team, we developed a multi-step process to manage, implement, and certify language translations to increase representation from study participants whose primary language is not English. This process increases data integrity by using written translations that captured data in a study participant’s language of choice and uses the same variable names across languages to facilitate data analysis.

## **MATERIALS AND METHODS**

The Northwestern University Data Analysis and Coordinating Center (NUACC) serves as the data coordinating center for multiple prospective clinical studies, managing all aspects of study

design, database development, data collection, data quality control, regulatory reporting, and statistical analyses. Given these studies enroll participants across geographically diverse regions throughout the United States, it is imperative to reduce barriers to enrollment for individuals whose native language may not be English. To ensure representative study samples, NUDACC manages the translation of study documents into languages other than English. Depending on the needs of the study, translated documents may include Informed Consent Forms, recruitment materials (e.g., brochures, advertisements, websites), and participant-facing CRFs. As part of ongoing internal process improvement efforts, NUDACC conducts periodic retrospective assessments of study operations where members of the team engage in open dialog about what went well, what could be improved, and adapt workflow for ongoing and future studies. This involves discussions amongst a collaborative study team with statistical, informatics, clinical research, regulatory, and project management expertise.

NUDACC's team members have been supporting studies requiring CRF translations for over a decade, and here we touch on several NUDACC-specific and NUDACC-affiliated studies that serve to illustrate how CRF translation processes have been implemented: QUARTET[10,11], Mothers and Babies[12], GO MOMs[13], and most recently the Liver Cirrhosis Network (LCN)[14]. These studies have provided opportunities to implement and refine our strategies for managing CRF language translations into a process that accounts for data structure and the intricacies of language in the clinical research context. Here we describe key study elements, system functionalities, and decisions that have helped to inform our proposed workflow, the NUDACC translation workflow (NTW).

### **Mode of Delivery**

Varying study needs and platform capabilities at the time of database development may require different approaches, for example: (1) creating duplicate electronic CRFs (eCRFs) within the same database for each language, (2) creating separate databases for each language, (3) modifying field labels to include translated text, or (4) using REDCap's[8,9] Multi-Language Management (MLM) module to translate eCRFs within the same project (Figure 1). These strategies have their own strengths and limitations for database creation, data collection, monitoring, and eventual analysis and are described further in Table 1 and in Supplemental

Methods. Regardless of the technical implementation, all methods require additional effort to ensure the translated material conveys the intended context while brokering cultural sensitivity. While the process we continue to describe here is centered on REDCap's MLM, there are other EDC systems with similar translation capabilities such as ClinInfo's ePro,[15] OpenEDC[16] and OpenClinica.[17]

## **RESULTS**

### **NUDACC Translation Workflow**

Leveraging the team's experience with translation of study materials as described above, we developed a workflow that provides an efficient and effective process for translating study documents; implementing translations into an electronic database; maintaining alignment between paper CRFs (pCRFs) and eCRFs; ensuring readability and comprehension in multiple languages; and facilitating data monitoring, reporting, and analysis (Figure 2).

NTW begins by creating an eCRF in the primary language (in our case, English) in REDCap, for any participant-facing forms. Once the eCRF is created, it is then duplicated in a word processing document to act as the pCRF, rather than printed directly from REDCap, to allow for explicit instructions and branching logic. We note non-technical challenges in this process. For instance, instructions may need to be modified to account for inherent differences between paper and electronic form instructions (e.g., "click" vs "circle", Supplemental Figure 1). This primary language pCRF is simultaneously submitted to the Institutional Review Board (IRB) and a translation contractor for certified translations into the target language(s). We also include a separate document of "supplemental translations" that may include eCRF-specific instructions such as page actions (e.g., "close", "cancel", "confirm") and site navigation (e.g., "next page", "previous page") that the end user may see.

Table 1. Description of example studies and multiple language display set up in REDCap.

Example Study (clinicaltrials.gov identifier)	Study Design	Languages	REDCap and Functionality**	Setup Pros	Cons
					Visually overwhelming for participant
<b>QUARTET USA</b>  (NCT03640312)  Start Year: 2019	Double-blind Randomized Controlled Trial (RCT)	English  Spanish	English and Spanish labels for questions and responses on the same CRF	No need to merge multiple fields on export  Less room for programming errors on export, especially for programmers unfamiliar with second language	Likely only possible with 2 total languages  Requires manual programming of field labels and code list options within REDCap
					Built-in buttons and commands cannot be translated (e.g., “next page”, “submit”)
<b>Mothers and Babies</b>  (NCT02979444)  Start Year: 2016	Cluster- randomized trial	English  Spanish	Spanish versions of assessments in a separate “arm” in REDCap; separate Spanish forms	Easier for coordination and data entry team to determine language and set of forms to send to participants  Translation of CRFs is “cleaner” and more efficient	Requires substantial programming and merging of fields on export  Opportunity for programming errors – both on merging data for programmers and in creating translated forms for those building database
					Built-in buttons and commands cannot be translated (e.g., “next page”, “submit”)
<b>GO MOMs</b>	Prospective	English	Separate Spanish	Same pros as “Mothers and	Same cons as “Mothers and Babies”

(NCT04860336)	observational study	Spanish	forms and data are backfilled into English forms via programming script nightly if filled out in Spanish	Babies” Study Backfilled Spanish to English makes monitoring possible for non-Spanish-speaking monitors	Study except, because each language has a separate CRF, it is not visually overwhelming to the participant.  Code to backfill Spanish to English takes time to generate and troubleshoot – added opportunity for programming errors
Start Year: 2021					
<b>Liver Cirrhosis Network</b>					
(NCT05740358; NCT05832229)	Prospective observational study	English	Multiple Language Management module	Easier for coordination team or participant to select the language they prefer for response	
		Spanish		No need to merge multiple fields on export	
		Simplified Chinese		Less room for programming errors on export, especially for programmers unfamiliar with second language	Requires substantial upfront programming and testing during database development
Start Year: 2022; 2023	Double-blind RCT			Built-in buttons and commands can be translated and added into end user interface (e.g., “next page”, “submit”)	
			Can efficiently handle >2 languages		

To utilize the REDCap MLM, described in detail in Supplemental Methods, we first create an initial MLM translation file for each language using a custom Python script shown in Supplemental Figure 2. We load each translation into REDCap, conducting a preliminary manual review and modification of any errors as needed. This script allows programmers who are unfamiliar with the language or alphabet used to successfully implement MLM, as long as the format and structure of the translation source documents remain consistent. The REDCap community also provides a shared library of translated phrases for the user interface—such as elements used in surveys or the survey submit buttons—and their certificate of translation.[18]

Once the translation is in place in REDCap, we asked native speakers (for LCN, this included both Spanish and Mandarin speakers) to review the translations using the MLM and/or pCRFs as our language contractor may not always have the research context to choose the best word or phrase. For instance, for the LCN study, we included a questionnaire about “food security”, which refers to how easily one feels they can access the food they need. Translators initially used a word that referred to the security of food quality (“is it safe to eat”), rather than the security of food access. Without the added step of native speaker review for the certified translations, the research-specific context would have been missed. Any suggestions are sent back to the translation contractor for review and recertification. Once the translations are finalized and the English versions are IRB-approved, we submit the translated pCRFs to the IRB with the certificate from the translation contractor. Finally, we distribute the pCRFs and sIRB approvals to sites, and release translations in REDCap MLM to production.

Once a study is up and running with translated participant materials, it is important to keep the translated participant materials synchronized with current participant materials. Any multi-centered trial will have a series of modifications, including adding or editing participant materials, throughout its lifecycle. During modifications, small changes may occur such as adding or removing text to clarify instructions. All changes undergo the entire process of (1) certified document translation, (2) native review, (3) recertification of translation with proposed native review changes; and (4) submission to IRB.



## **Impact**

Using NTW, LCN has translated 15 CRFs into two languages, and are enrolling participants who complete surveys in English, Spanish, and Simplified Chinese. During development, native speakers were able to identify multiple errors in both the original translation and REDCap programming, which highlights the need for this step in NTW. Contrary to other options presented in Table 1, study statisticians can quickly and easily query data completion and quality by language to spot any trends or issues in translation as the study progresses, and the burden of data restructuring by language is eliminated. Recruitment is ongoing and participants in all three languages have been enrolled.

## **DISCUSSION**

REDCap and other EDC systems support multiple ways to implement translated eCRFs. Here we describe a process that includes several considerations beyond the technical implementation that are critical to ensure similar engagement and comprehension from non-English speaking participants as compared to native English speakers. NTW requires additional time and budget, native-speaking personnel, and regulatory expertise as well as complex document management are crucial. Implementing NTW requires the largest up-front investment with the first set of translations; however, the workload diminishes greatly once the materials are finalized.

Our translation process navigates some of the issues posed by other research teams such as Colina et al., including readability, comprehension, and data validity.[7,19] For studies unable to undertake full validation, cultural adaptation, and a functionalist approach, we proposed a straightforward alternative that incorporates review by native speakers who can provide both cultural heritage and the research knowledge to avoid the common mistakes of translating, not translating, or providing out of context study materials.

One challenge we encountered was accommodating regional dialects – for example, several studies include Spanish-speaking participants living in Florida and Southern California who speak different dialects. This required occasional discussion amongst the translation service and our native speakers to agree on appropriate phrasing. We also acknowledge the added care needed for translating validated questionnaires; in this case close collaboration with the study

investigators identified existing validations and copyright issues. Additionally, appropriate reading level of translated text can be easily overlooked when translating study materials if not explicitly indicated.

We acknowledge this work has several limitations. NTW does not follow a cultural adaptation process, does not produce validated research instruments (though validated instruments can be incorporated into the MLM in REDCap, benefitting from the same data structure advantages), and we did not incorporate a focus group or review by our participant population. The pool of native speakers reviewing the translation in LCN were not representative of all Spanish-speaking regions, for example Spain, which often diverts from other dialects in both word choice and pronunciation. As data collection for LCN is ongoing, we were not able to directly compare data quality between NTW and previous methods. Last, while we believe NTW may be more cost-effective than other alternatives, future work should include a formal cost comparison study.

## CONCLUSION

The multi-step translation process for CRFs proposed here offers an adaptable approach that can be implemented using a variety of EDC systems. NTW allows for increased representation in clinical study populations, improved understanding of questions and responses provided in a participant's primary language, a streamlined alternative for investigator-initiated studies, and more straightforward data processing.

## AUTHOR CONTRIBUTIONS

**Amy Krefman:** Conceptualization, Project administration, Visualization, Writing-original draft, Writing-review & editing; **Luke Rasmussen:** Conceptualization, Software, Visualization, Writing-review & editing; **Crystal Santillanes:** Conceptualization, Investigation, Writing-review & editing; **Mary Beth Tull:** Conceptualization, Writing-original draft, Writing-review & editing; **Patricia Bustamante:** Writing-review & editing; **Alan Kuang:** Writing-original draft, Writing-review & editing; **Hongyan Ning:** Investigation, Validation, Writing-review & editing; **Yaojie Wang:** Investigation, Validation, Writing-review & editing; **Barbara Gonzalez:** Investigation, Writing-review & editing; **Eugenia Manrique:** Investigation, Writing-review & editing; **Lauren**

**Bonner:** Supervision, Writing-review & editing; **Jody Ciolino:** Conceptualization, Funding acquisition, Supervision, Visualization, Writing-review & editing.

## **ACKNOWLEDGEMENTS**

We wish to thank Rachel Moore and Elizabeth Yan for their input on translated material workflows.

## **COMPETING INTERESTS**

The authors have no competing interests to declare.

## **FUNDING**

This research is supported by NIDDK, NIAAA, and NCI via the following grant: U24DK130164, Northwestern University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

NUCATS is funded in part by a Clinical and Translational Science Award (CTSA) grant from the National Institutes of Health (NIH). UL1TR001422

## **DATA AVAILABILITY STATEMENT**

No new data were generated or analyzed in support of this research.

## Figure 1. Overview of REDCap Setup to Accommodate eCRF Translations across Multiple Studies

(a) In this example demographics form from the QUARTET USA study, both English and Spanish labels for questions and responses are included for any one field. Upon export (the back end), the programmer would need to only use one set of fields for data merging, restructuring and analyses; however, with certain fields and responses, this setup may be visually overwhelming for the data enterer or study participant taking the survey. (b) The Mothers and Babies study duplicated all surveys (participant-facing eCRFs) and used the “arm” feature in REDCap to delineate between participants taking the surveys in English (depicted here) and those that are taking the surveys in Spanish (in Arms 3 and 4, with the corresponding Spanish forms linked to the appropriate events). (c) In both the Mothers and Babies and the GO MOMs study examples this requires duplicate sets of forms and thus duplicate sets of fields, resulting in more programming and merging on the back end prior to analyses. (d) The Multiple Language Management Module (MLM) allows the data enterer/study participant to toggle to their preferred language in real-time. Upon export, the programmer will only have one set of fields to use.

(a) QUARTET USA Example Demographics CRF

Visit Date  
Fecha de la visita

Date of Birth  
Fecha de nacimiento

Age  
Edad

Sex as defined at birth  
Sexo definido al nacer

(b) Mothers and Babies REDCap Setup

Arm 1: Intervention Arm 2: Control Arm 3: SP-Intervention Arm 4: SP-Control Arm

Arm name: Intervention

Begin Editing

Data Collection Instrument	Baseline	Post	12 Weeks	24 Weeks
EPDS (locked)	↓			
EPDS Spanish (locked)	↓			
Involvement in Prenatal Services (locked)	↓	↓	↓	↓
Involvement in Prenatal Services Spanish (locked)	↓	↓	↓	↓
QIDS-SR16 (locked)	↓	↓	↓	↓
QIDS-SR16 Spanish (locked)	↓	↓	↓	↓
QIDS Post Closing (locked)		✓		
QIDS Post Closing Spanish (locked)		✓		
Social Support Survey (locked)	↓	↓	↓	↓
Social Support Survey Spanish (locked)	↓	↓	↓	↓
Maternal Mood Screener (locked)	↓	↓	↓	↓
Maternal Mood Screener Spanish (locked)	↓	↓	↓	↓

(c) Duplicate Forms / Fields as in Mothers and Babies and GO MOMs

Demographics

1. What is your date of birth?

2. What is your current marital status?

Demographics

1. ¿Cuál es su fecha de nacimiento?

2. ¿Cuál es su estado civil actual?

(d) Multiple Language Management Module as in LCN

Demographics

English Spanish Chinese

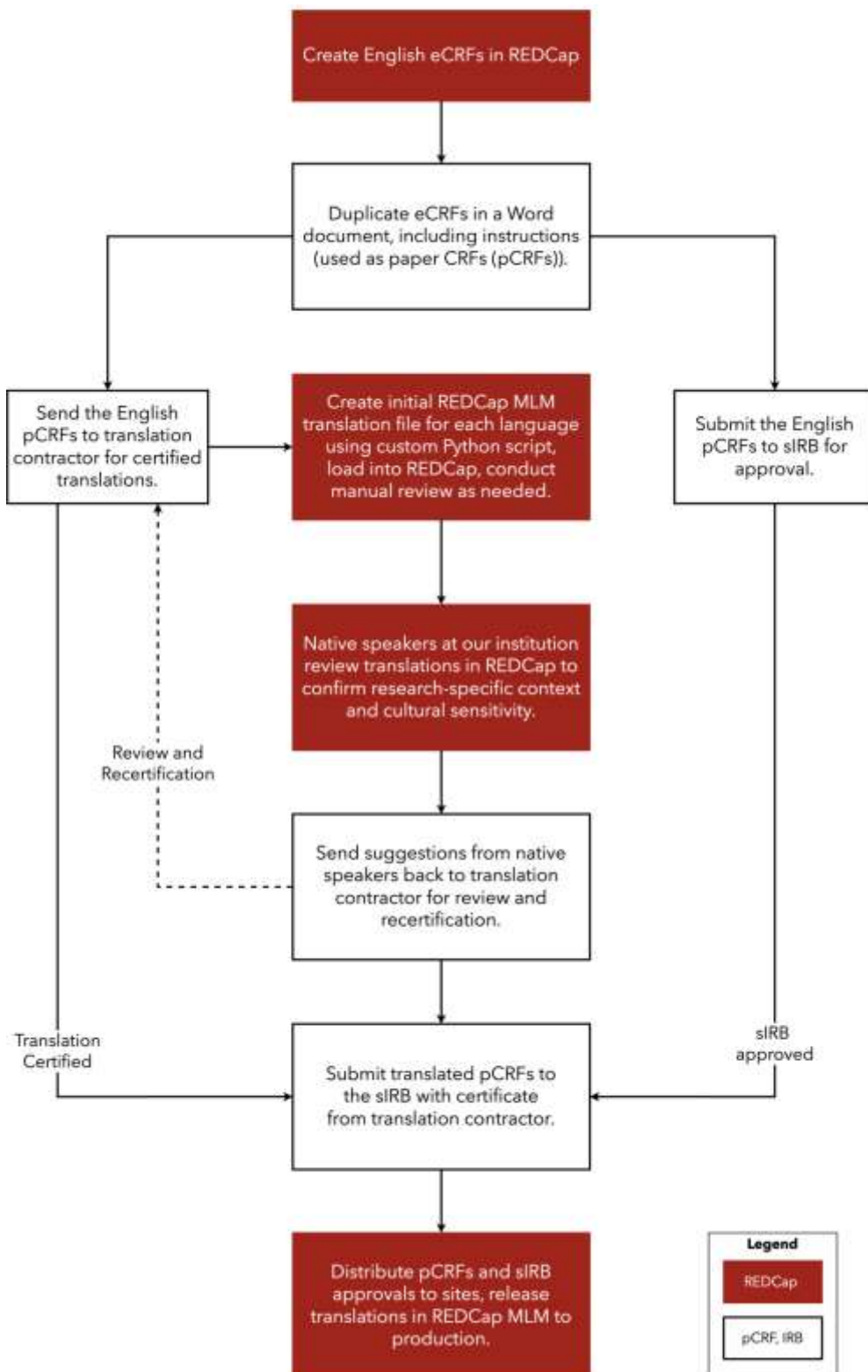
We would like to collect some basic information about your background. This is important in any research study because it helps the public to understand if research results are applicable to them.

Today's date: 08-05-2024

Age: 48.0

Sex as assigned at birth: Female

**Figure 2. Flow chart describing the NUDACC Translation Workflow process**



## REFERENCES

1. National Institute on Minority Health and Health Disparities. Diversity and Inclusion in Clinical Trials. Updated 12 January 2024. Accessed 25 July 2024, <https://www.nimhd.nih.gov/resources/understanding-health-disparities/diversity-and-inclusion-in-clinical-trials.html>
2. Peloquin D, Levine GH, Barnes M. FDA Finalizes Guidance on Informed Consent for Clinical Investigations. Ropes & Gray. Updated 24 August 2023. <https://www.ropesgray.com/en/insights/alerts/2023/08/fda-finalizes-guidance-on-informed-consent-for-clinical-investigations>
3. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims (2009).
4. DuBay M, Sideris J, Rouch E. Is traditional back translation enough? Comparison of translation methodology for an ASD screening tool. *Autism Res.* Oct 2022;15(10):1868-1882. doi:10.1002/aur.2783
5. Jiang L, Lv M, Wen Y, Zhang P, Huang Q. Bilingual Conversion in the Translation of ICH Terms: A Study on the Psychological Processes of Translators. *J Psycholinguist Res.* Dec 2023;52(6):2239-2256. doi:10.1007/s10936-023-09989-6
6. Velez MA, Glenn BA, Garcia-Jimenez M, et al. Consent document translation expense hinders inclusive clinical trial enrolment. *Nature.* Aug 2023;620(7975):855-862. doi:10.1038/s41586-023-06382-0
7. Colina S, Rodriguez-Guerra M, Marrone N, et al. Research Documents for Populations with Limited English Proficiency: Translation Approaches Matter. *Ethics Hum Res.* Jan 2022;44(1):29-39. doi:10.1002/eahr.500115
8. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform.* Jul 2019;95:103208. doi:10.1016/j.jbi.2019.103208
9. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* Apr 2009;42(2):377-81. doi:10.1016/j.jbi.2008.08.010

10. Baldrige AS, Huffman MD, Lazar D, et al. Efficacy and safety of a quadruple ultra-low-dose treatment for hypertension (QUARTET USA): Rationale and design for a randomized controlled trial. *Am Heart J*. Dec 2022;254:183-193. doi:10.1016/j.ahj.2022.09.004
11. Huffman MD, Baldrige AS, Lazar D, et al. Efficacy and safety of a four-drug, quarter-dose treatment for hypertension: the QUARTET USA randomized trial. *Hypertens Res*. Jun 2024;47(6):1668-1677. doi:10.1038/s41440-024-01658-y
12. Jensen JK, Ciolino JD, Diebold A, et al. Comparing the Effectiveness of Clinicians and Paraprofessionals to Reduce Disparities in Perinatal Depression via the Mothers and Babies Course: Protocol for a Cluster-Randomized Controlled Trial. Protocol. *JMIR Res Protoc*. 2018;7(11):e11624. doi:10.2196/11624
13. Group GOMS. Design, rationale and protocol for Glycemic Observation and Metabolic Outcomes in Mothers and Offspring (GO MOMs): an observational cohort study. *BMJ Open*. Jun 8 2024;14(6):e084216. doi:10.1136/bmjopen-2024-084216
14. Liver Cirrhosis Network. Northwestern University. Accessed 25 July 2024, <https://www.lcnstudy.org/>
15. ClinInfo: Custom E-Clinical Solutions. Accessed 2 July 2024, <http://www.clininfo.fr/applications/ePRO>
16. Greulich L, Hagselmann S, Dugas M. An Open-Source, Standard-Compliant, and Mobile Electronic Data Capture System for Medical Research (OpenEDC): Design and Evaluation Study. *JMIR Med Inform*. Nov 19 2021;9(11):e29176. doi:10.2196/29176
17. Tamm K. OpenClinica Documentation and Training Site. Updated 11 September 2020. Accessed 2 July 2024, <https://docs.openclinica.com/oc4/>
18. Multi-Language Management (MLM) Translations Repository. REDCap. Accessed 15 July 2024, [https://redcap.vumc.org/surveys/?\\_report=7LEEXCTLYA7TJDJ3](https://redcap.vumc.org/surveys/?_report=7LEEXCTLYA7TJDJ3)
19. Colina S, Marrone N, Ingram M, Sanchez D. Translation Quality Assessment in Health Research: A Functionalist Alternative to Back-Translation. *Eval Health Prof*. Sep 2017;40(3):267-293. doi:10.1177/0163278716648191