Φysikotherapia
Volume 26 Issue 4 October - December 2023

The official Greek version of the National Institutes of Health - Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies

Stasi Sophia1, Akermanidis Theologos1, Karamouza Paraskevi1, Stasinopoulos Dimitrios1

1 Laboratory of Neuromuscular and Cardiovascular Study of Motion (LANECASM), Department of Physiotherapy, University of West Attica (UNIWA)

DOI: 10.55742/BJSA8943

ABSTRACT

Purpose: The aim of this study was the official translation into Greek of the "National Institutes of Health - Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies" (NIH-QATOCCs) proposed by the American National Institutes of Health.

Materials - Method: Official permission for translating the original NIH-QATOCCsS questionnaire and its guidance was given by the American National Heart, Lung, and Blood Institute (NHLBI). The adaptation of NIH-QATOCCsS into Greek followed Brislin's classic back-translation model, which included: (a) back-translation method, (b) bilingual technique, (c) committee approach, and (d) pretest procedure.

Results: After completing all the required procedures, the final Greek version of the NIH-QATOCCsS was approved by the present study's authors.

Conclusions: The NIH-QATOCCsS is a detailed tool, easy to use, with clear guidance, and its Greek translation is available for scientists and students writing in Greek, systematic reviews that include observational cohort or cross-sectional studies.

Key words: Observational study, Methodological quality, Assessment, Systematic review

Introduction

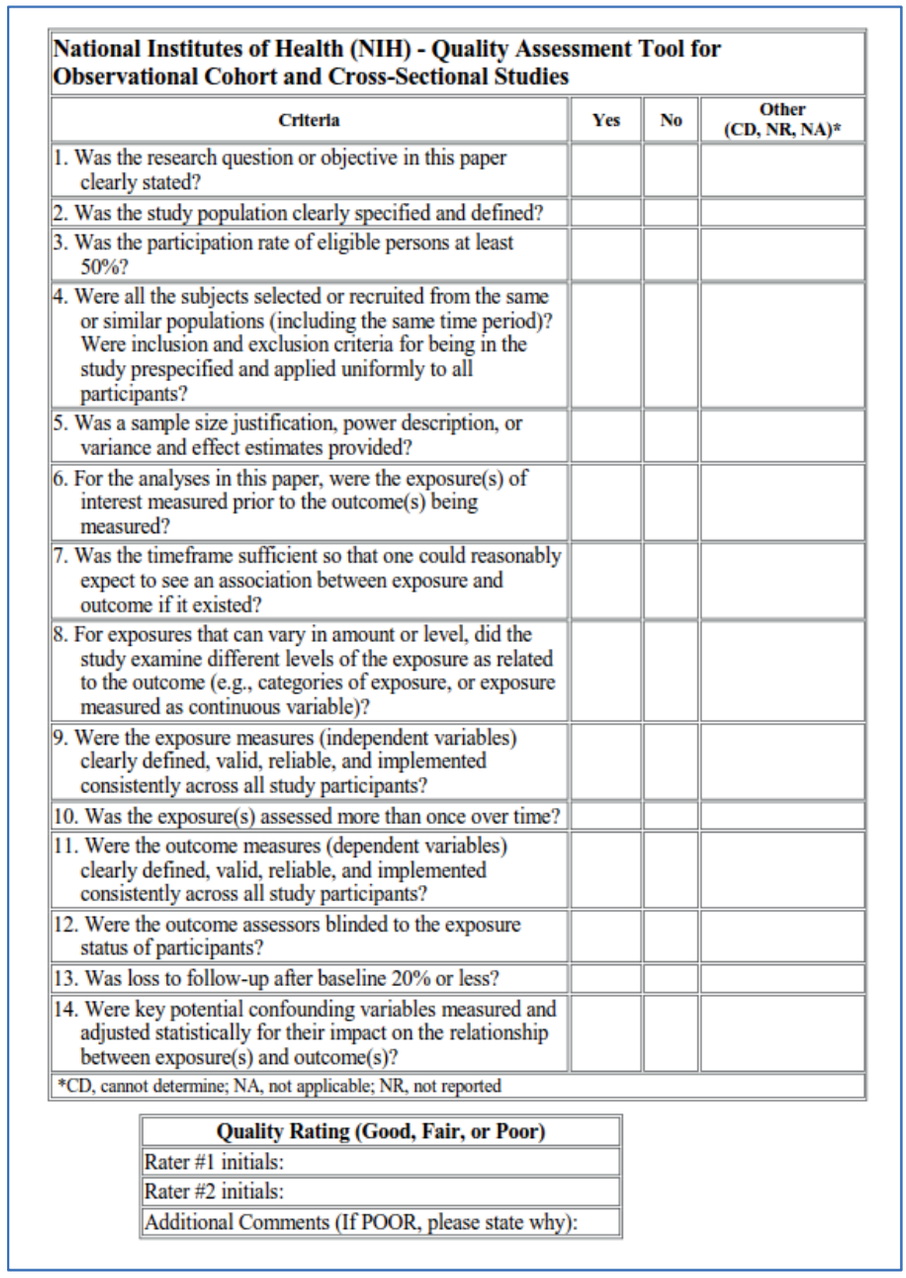
In the health sciences, systematic reviews are valuable for acquiring and exchanging knowledge as they summarize and analyze findings from individual studies (Drukker et al, 2021). The evaluation of the methodological quality of the included studies is considered necessary for the assessment of the internal validity of a study (Drukker et al, 2021). Various methodological quality assessment tools are available for non- or randomized controlled trials, case studies, and observational studies (Drukker et al, 2021). However, they are written in English, and a possible misunderstanding of the tool's questions by a non-native English author poses a risk of misinterpretation. Generally, direct translation of an instrument from one language to another does not guarantee content equivalence of the translated scale (Brislin, 1970; Sechrest & Fay, 1972). Researchers agree that the back-translation of an instrument is essential for its validation (Brislin, 1970; Jones et al., 2001).

The purpose of this study was the official translation into Greek of the “National Institutes of Health - Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies” (NIH-QATOCCsS) of the American National Institutes of Health.

Materials – Method

National Institutes of Health - Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies

The NIH - QATOCCsS contains 14 questions that assess the internal validity and risk of potential bias of the parameters: sample selection, information, measurements, or confounding factors on the effects of exposures on the outcomes/results of the included studies (Figure 1). The methodological evaluation is carried out by two assessors, and the possible answers to each question are: "Yes", "No", or "Other" (Not specified/ Not applicable/ Not mentioned). After the 14 questions are answered, the assessor rates the quality of the research as "Good" ("Yes" to 11–14 out of the 14 questions), "Moderate" ("Yes" to 5–10 out of the 14 questions), or "Poor "("Yes" to 0–4 out of the 14 questions). If a study is rated as "Poor" the reviewers explain why (NHLBI, 2021; Bagias et al, 2021).



National Institutes of Health (NIH) - Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies

Criteria	Yes	No	Other (CD, NR, NA)*
1. Was the research question or objective in this paper clearly stated?			
2. Was the study population clearly specified and defined?			
3. Was the participation rate of eligible persons at least 50%?			
4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?			
5. Was a sample size justification, power description, or variance and effect estimates provided?			
6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?			
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?			
8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)?			
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
10. Was the exposure(s) assessed more than once over time?			
11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
12. Were the outcome assessors blinded to the exposure status of participants?			
13. Was loss to follow-up after baseline 20% or less?			
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?			

*CD, cannot determine; NA, not applicable; NR, not reported

Quality Rating (Good, Fair, or Poor)
Rater #1 initials:
Rater #2 initials:
Additional Comments (If POOR, please state why):

Figure 1. The original NIH - Quality Assessment Tool questionnaire for Observational Cohort and Cross-Sectional Studies.

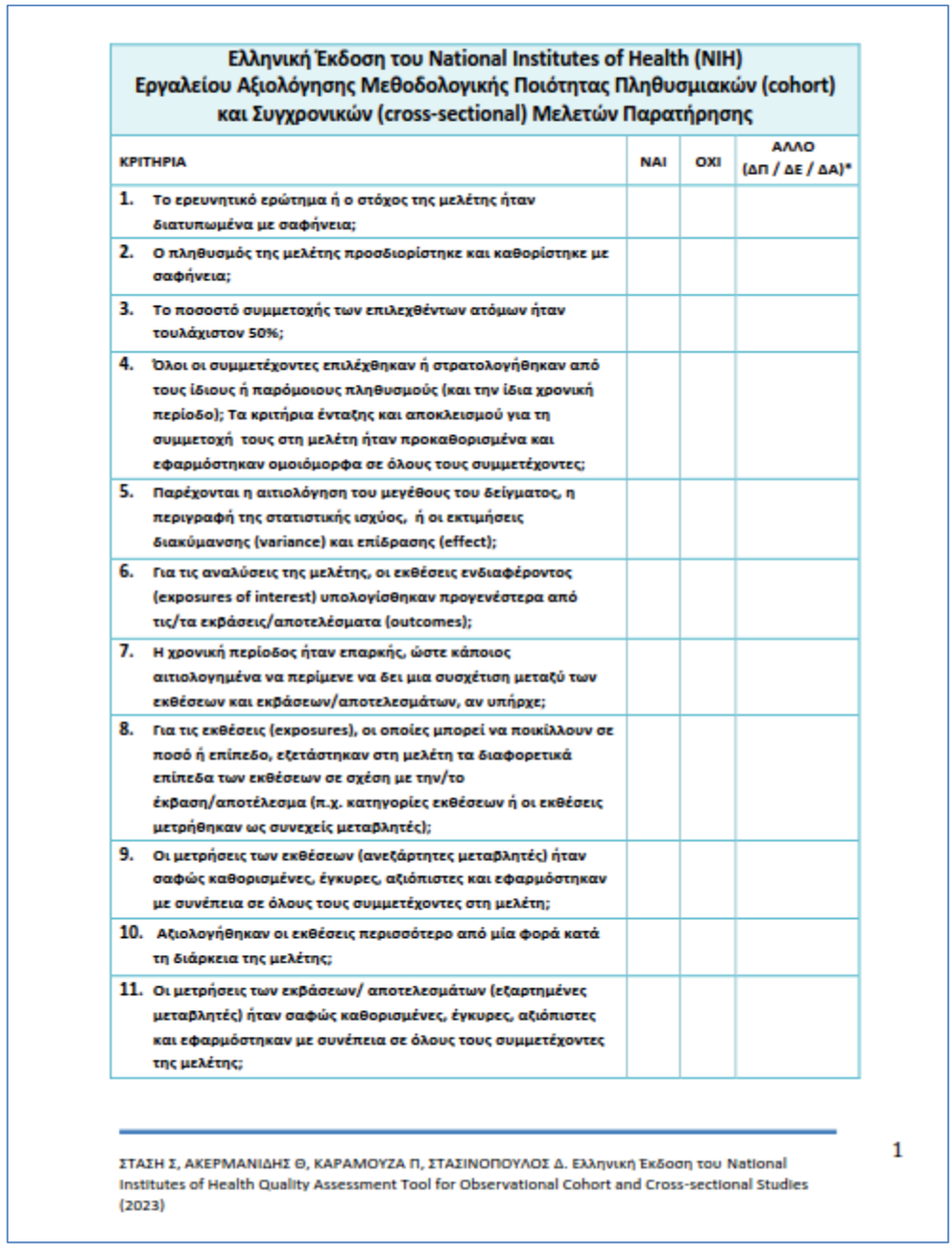
Translation Procedures

Official permission for translating the original NIH- QATOCCsS questionnaire and its guidances was given by the American National Heart, Lung, and Blood Institute (NHLBI).

The adaptation of NIH- QATOCCsS into Greek followed the guidelines proposed by Brislin's classic back-translation model (Brislin, 1970; Brislin, 1976) and Jones et al. (2001). Brislin (1970) offered four techniques for maintaining the equivalence between the original and translated measures: (a) back-translation method, (b) bilingual technique, (c) committee approach, and (d) pretest procedure.

According to the aforementioned model, a bilingual non-medical specialist blindly translated the questionnaire and the guidance of the NIH- QATOCCsS from English to

Greek; a second bilingual translator independently back-translated the Greek documents back to English. Next, the two English versions of the instrument (original and back-translated versions) were compared for concept equivalence. Technical and linguistic adaptations were agreed upon in a consensus meeting. During the “committee approach” phase, a team of experts consisting of physiotherapists and three bilingual non-medical specialists took care of all the required procedures. Finally, according to Brislin's pretest procedure, a comprehension test of the NIH- QATOCCsS's Greek version was carried out by a group of health scientists (N=15) unfamiliar with the English version.



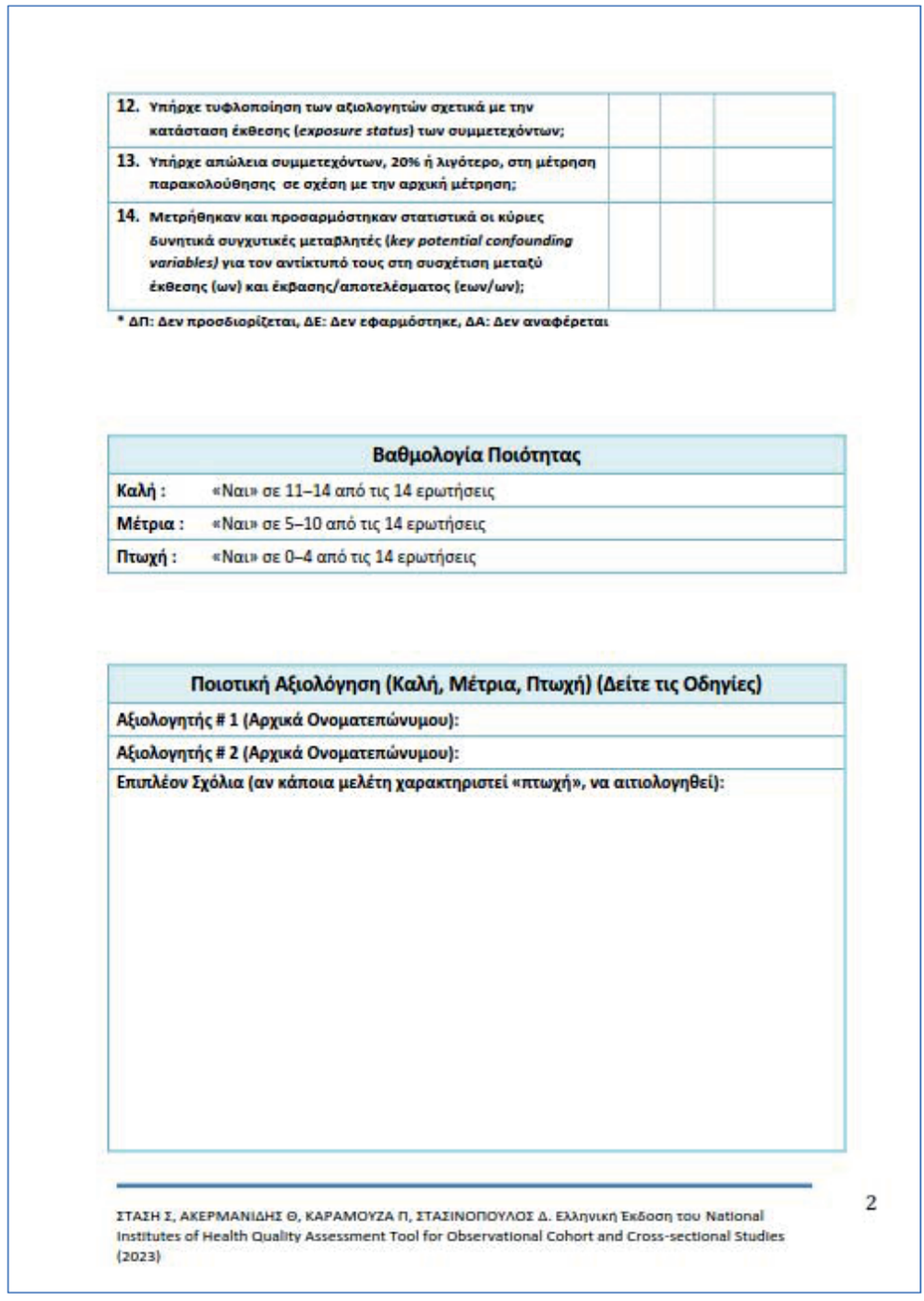
Ελληνική Έκδοση του National Institutes of Health (NIH) Εργαλείου Αξιολόγησης Μεθοδολογικής Ποιότητας Πληθυσμιακών (cohort) και Συγχρονικών (cross-sectional) Μελετών Παρατήρησης

ΚΡΙΤΗΡΙΑ	ΝΑΙ	ΟΧΙ	ΑΛΛΟ (ΔΠ / ΔΕ / ΔΑ)*
1. Το ερευνητικό ερώτημα ή ο στόχος της μελέτης ήταν διατυπωμένα με σαφήνεια;			
2. Ο πληθυσμός της μελέτης προσδιορίστηκε και καθορίστηκε με σαφήνεια;			
3. Το ποσοστό συμμετοχής των επιλεχθέντων ατόμων ήταν τουλάχιστον 50%;			
4. Όλοι οι συμμετέχοντες επιλέχθηκαν ή στρατολογήθηκαν από τους ίδιους ή παρόμοιους πληθυσμούς (και την ίδια χρονική περίοδο); Τα κριτήρια ένταξης και αποκλεισμού για τη συμμετοχή τους στη μελέτη ήταν προκαθορισμένα και εφαρμόστηκαν ομοιόμορφα σε όλους τους συμμετέχοντες;			
5. Παρέχονται η αιτιολόγηση του μεγέθους του δείγματος, η περιγραφή της στατιστικής ισχύος, ή οι εκτιμήσεις διακύμανσης (variance) και επίδρασης (effect);			
6. Για τις αναλύσεις της μελέτης, οι εκθέσεις ενδιαφέροντος (exposures of interest) υπολογίσθηκαν προγενέστερα από τις/τα εκβάσεις/αποτελέσματα (outcomes);			
7. Η χρονική περίοδος ήταν επαρκής, ώστε κάποιος αιτιολογημένα να περίμενε να δει μια συσχέτιση μεταξύ των εκθέσεων και εκβάσεων/αποτελεσμάτων, αν υπήρχε;			
8. Για τις εκθέσεις (exposures), οι οποίες μπορεί να ποικίλλουν σε ποσό ή επίπεδο, εξετάστηκαν στη μελέτη τα διαφορετικά επίπεδα των εκθέσεων σε σχέση με την/το έκβαση/αποτέλεσμα (π.χ. κατηγορίες εκθέσεων ή οι εκθέσεις μετρήθηκαν ως συνεχείς μεταβλητές);			
9. Οι μετρήσεις των εκθέσεων (ανεξάρτητες μεταβλητές) ήταν σαφώς καθορισμένες, έγκυρες, αξιόπιστες και εφαρμόστηκαν με συνέπεια σε όλους τους συμμετέχοντες στη μελέτη;			
10. Αξιολογήθηκαν οι εκθέσεις περισσότερο από μία φορά κατά τη διάρκεια της μελέτης;			
11. Οι μετρήσεις των εκβάσεων/ αποτελεσμάτων (εξαρτημένες μεταβλητές) ήταν σαφώς καθορισμένες, έγκυρες, αξιόπιστες και εφαρμόστηκαν με συνέπεια σε όλους τους συμμετέχοντες της μελέτης;			

ΣΤΑΣΗ Σ, ΑΚΕΡΜΑΝΙΔΗΣ Θ, ΚΑΡΑΜΟΥΖΑ Π, ΣΤΑΣΙΝΟΠΟΥΛΟΣ Δ. Ελληνική Έκδοση του National Institutes of Health Quality Assessment Tool for Observational Cohort and Cross-sectional Studies (2023)

1

Figure 2: The 1st page of the Greek questionnaire of the NIH - Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies.



12. Υπήρχε τυφλοποίηση των αξιολογητών σχετικά με την κατάσταση έκθεσης (*exposure status*) των συμμετεχόντων;			
13. Υπήρχε απώλεια συμμετεχόντων, 20% ή λιγότερο, στη μέτρηση παρακολούθησης σε σχέση με την αρχική μέτρηση;			
14. Μετρήθηκαν και προσαρμόστηκαν στατιστικά οι κύριες δυνητικά συγχυτικές μεταβλητές (*key potential confounding variables*) για τον αντίκτυπό τους στη συσχέτιση μεταξύ έκθεσης (ων) και έκβασης/αποτελέσματος (εων/ων);			

* ΔΠ: Δεν προσδιορίζεται, ΔΕ: Δεν εφαρμόστηκε, ΔΑ: Δεν αναφέρεται

Βαθμολογία Ποιότητας	
Καλή :	«Ναι» σε 11–14 από τις 14 ερωτήσεις
Μέτρια :	«Ναι» σε 5–10 από τις 14 ερωτήσεις
Πτωχή :	«Ναι» σε 0–4 από τις 14 ερωτήσεις

Ποιοτική Αξιολόγηση (Καλή, Μέτρια, Πτωχή) (Δείτε τις Οδηγίες)
Αξιολογητής # 1 (Αρχικά Ονοματεπώνυμου):
Αξιολογητής # 2 (Αρχικά Ονοματεπώνυμου):
Επιπλέον Σχόλια (αν κάποια μελέτη χαρακτηριστεί «πτωχή», να αιτιολογηθεί):

2

ΣΤΑΣΗ Σ, ΑΚΕΡΜΑΝΙΔΗΣ Θ, ΚΑΡΑΜΟΥΖΑ Π, ΣΤΑΣΙΝΟΠΟΥΛΟΣ Δ. Ελληνική Έκδοση του National Institutes of Health Quality Assessment Tool for Observational Cohort and Cross-sectional Studies (2023)

Figure 3: The 2nd page of the Greek questionnaire of the NIH - Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies.

Results

During Brislin's “committee approach” phase, the phrase "exposure of interest" was translated as "exposure to the factors under study" and the word "outcome" as "outcomes/results". The vast majority of health scientists involved to the Greek questionnaire's comprehension test understood these phrases. The final Greek version of the NIH- QATOCCsS is presented in Figures 2 & 3.

The detailed guidance for assessing the quality of observational cohort and cross-sectional studies is documented below. It is organized by question number.

Question 1. Research question

Did the authors describe their goal in conducting this research? Is it easy to understand what they were looking to find? This issue is important for any scientific paper of any type. Higher-quality scientific research explicitly defines a research question (NHLBI, 2021).

Questions 2 and 3. Study population

Did the authors describe the group of people from which the study participants were selected or recruited, using demographics, location, and time period? If you were to conduct this study again, would you know who to recruit, from where, and from what time period? Is the cohort population free of the outcomes of interest at the time they were recruited?

An example would be men over 40 years old with type 2 diabetes who began seeking medical care at Phoenix Good Samaritan Hospital between January 1, 1990 and December 31, 1994. In this example, the population is clearly described as: (1) who (men over 40 years old with type 2 diabetes); (2) where (Phoenix Good Samaritan Hospital); and (3) when (between January 1, 1990 and December 31, 1994). Another example is women ages 34 to 59 years of age in 1980 who were in the nursing profession and had no known coronary disease, stroke, cancer, hypercholesterolemia, or diabetes, and were recruited from the 11 most populous States, with contact information obtained from State nursing boards.

In cohort studies, it is crucial that the population at baseline is free of the outcome of interest. For example, the nurses' population above would be an appropriate group in which to study incident coronary disease. This information is usually found either in descriptions of population recruitment, definitions of variables, or inclusion/exclusion criteria.

Reviewer may need to look at prior papers on methods in order to make the assessment for this question. Those papers are usually in the reference list. If fewer than 50% of eligible persons participated in the study, then there is concern that the study population does not adequately represent the target population. This increases the risk of bias (NHLBI, 2021).

Question 4. Groups recruited from the same population and uniform eligibility criteria

Were the inclusion and exclusion criteria developed prior to recruitment or selection of the study population? Were the same underlying criteria used for all of the subjects involved? This issue is related to the description of the study population, above, and you may find the information for both of these questions in the same section of the paper.

Most cohort studies begin with the selection of the cohort; participants in this cohort are then measured or evaluated to determine their exposure status. However, some cohort studies may recruit or select exposed participants in a different time or place than unexposed participants, especially retrospective cohort studies–which is when data are obtained from the

past (retrospectively), but the analysis examines exposures prior to outcomes. For example, one research question could be whether diabetic men with clinical depression are at higher risk for cardiovascular disease than those without clinical depression. So, diabetic men with depression might be selected from a mental health clinic, while diabetic men without depression might be selected from an internal medicine or endocrinology clinic. This study recruits groups from different clinic populations, so this example would get a "no."

However, the women nurses described in the question above were selected based on the same inclusion/exclusion criteria, so that example would get a "yes" (NHLBI, 2021).

Question 5. Sample size justification

Did the authors present their reasons for selecting or recruiting the number of people included or analyzed? Do they note or discuss the statistical power of the study? This question is about whether or not the study had enough participants to detect an association if one truly existed.

A paragraph in the article's methods section may explain the sample size needed to detect a hypothesized difference in outcomes. You may also find a discussion of power in the discussion section (such as the study had 85 percent power to detect a 20 percent increase in the rate of an outcome of interest, with a 2-sided alpha of 0.05). Sometimes estimates of variance and/or estimates of effect size are given, instead of sample size calculations. In any of these cases, the answer would be "yes."

However, observational cohort studies often do not report anything about power or sample sizes because the analyses are exploratory in nature. In this case, the answer would be "no." This is not a "fatal flaw." It just may indicate that attention was not paid to whether the study was sufficiently sized to answer a prespecified question–i.e., it may have been an exploratory, hypothesis-generating study (NHLBI, 2021).

Question 6. Exposure assessed prior to outcome measurement

This question is important because, in order to determine whether an exposure causes an outcome, the exposure must come before the outcome.

For some prospective cohort studies, the investigator enrolls the cohort and then determines the exposure status of various members of the cohort (large epidemiological studies like Framingham used this approach). However, for other cohort studies, the cohort is selected based on its exposure status, as in the example above of depressed diabetic men (the exposure being depression). Other examples include a cohort identified by its exposure to fluoridated drinking water and then compared to a cohort living in an area without fluoridated water, or a cohort of military personnel exposed to combat in the Gulf War compared to a cohort of military personnel not deployed in a combat zone.

With either of these types of cohort studies, the cohort is followed forward in time (i.e., prospectively) to assess the outcomes that occurred in the exposed members compared to nonexposed members of the cohort. Therefore, you begin the study in the present by looking at groups that were exposed (or not) to some biological or behavioral factor, intervention, etc., and then you follow them forward in time to examine outcomes. If a cohort study is conducted properly, the answer to this question should be "yes," since the exposure status of members of the cohort was determined at the beginning of the study before the outcomes occurred.

For retrospective cohort studies, the same principle applies. The difference is that, rather than identifying a cohort in the present and following them forward in time, the investigators go back in time (i.e., retrospectively) and select a cohort based on their exposure status in the past and then follow them forward to assess the outcomes that occurred in the exposed and nonexposed cohort members. Because in retrospective cohort studies the exposure and outcomes may have already occurred (it depends on how long they follow the cohort), it is important to make sure that the exposure preceded the outcome.

Sometimes cross-sectional studies are conducted (or cross-sectional analyses of cohort-study data), where the exposures and outcomes are measured during the same timeframe. As a result, cross-sectional analyses provide weaker evidence than regular cohort studies regarding a potential causal relationship between exposures and outcomes. For cross-sectional analyses, the answer to Question 6 should be “no” (NHLBI, 2021).

Question 7. Sufficient timeframe to see an effect

Did the study allow enough time for a sufficient number of outcomes to occur or be observed, or enough time for an exposure to have a biological effect on an outcome? In the examples given above, if clinical depression has a biological effect on increasing risk for CVD, such an effect may take years. In the other example, if higher dietary sodium increases BP, a short timeframe may be sufficient to assess its association with BP, but a longer timeframe would be needed to examine its association with heart attacks.

The issue of timeframe is important to enable meaningful analysis of the relationships between exposures and outcomes to be conducted. This often requires at least several years, especially when looking at health outcomes, but it depends on the research question and outcomes being examined.

Cross-sectional analyses allow no time to see an effect, since the exposures and outcomes are assessed at the same time, so those would get a "no" response (NHLBI, 2021).

Question 8. Different levels of the exposure of interest

If the exposure can be defined as a range (examples: drug dosage, amount of physical activity, amount of sodium consumed), were multiple categories of that exposure assessed? (for example, for drugs: not on the medication, on a low dose, medium dose, high dose; for dietary sodium, higher than average U.S. consumption, lower than recommended consumption, between the two). Sometimes discrete categories of exposure are not used, but instead exposures are measured as continuous variables (for example, mg/day of dietary sodium or BP values).

In any case, studying different levels of exposure (where possible) enables investigators to assess trends or dose-response relationships between exposures and outcomes–e.g., the higher the exposure, the greater the rate of the health outcome. The presence of trends or dose-response relationships lends credibility to the hypothesis of causality between exposure and outcome.

For some exposures, however, this question may not be applicable (e.g., the exposure may be a dichotomous variable like living in a rural setting versus an urban setting, or vaccinated/not vaccinated with a one-time vaccine). If there are only two possible exposures (yes/no), then this question should be given an "NA," and it should not count negatively towards the quality rating (NHLBI, 2021).

Question 9. Exposure measures and assessment

Were the exposure measures defined in detail? Were the tools or methods used to measure exposure accurate and reliable–for example, have they been validated or are they objective? This issue is important as it influences confidence in the reported exposures. When exposures are measured with less accuracy or validity, it is harder to see an association between exposure and outcome even if one exists. Also as important is whether the exposures were assessed in the same manner within groups and between groups; if not, bias may result.

For example, retrospective self-report of dietary salt intake is not as valid and reliable as prospectively using a standardized dietary log plus testing participants' urine for sodium content. Another example is measurement of BP, where there may be quite a difference between usual care, where clinicians measure BP however it is done in their practice setting (which can vary considerably), and use of trained BP assessors using standardized equipment (e.g., the same BP device which has been tested and calibrated) and a standardized protocol (e.g., patient is seated for 5 minutes with feet flat on the floor, BP is taken twice in each arm, and all four measurements are averaged). In each of these cases, the former would get a "no" and the latter a “yes”.

Here is a final example that illustrates the point about why it is important to assess exposures consistently across all groups: If people with higher BP (exposed cohort) are seen by their providers more frequently than those without elevated BP (nonexposed group), it also increases the chances of detecting and documenting changes in health outcomes, including CVD-related events. Therefore, it may lead to the conclusion that higher BP leads to more CVD events. This may be true, but it could also be due to the fact that the subjects with higher BP were seen more often; thus, more CVD-related events were detected and documented simply because they had more encounters with the health care system. Thus, it could bias the results and lead to an erroneous conclusion (NHLBI, 2021).

Question 10. Repeated exposure assessment

Was the exposure for each person measured more than once during the course of the study period? Multiple measurements with the same result increase our confidence that the exposure status was correctly classified. Also, multiple measurements enable investigators to look at changes in exposure over time, for example, people who ate high dietary sodium throughout the followup period, compared to those who started out high and then reduced their intake, compared to those who ate low sodium throughout. Once again, this may not be applicable in all cases. In many older studies, exposure was measured only at baseline. However, multiple exposure measurements do result in a stronger study design (NHLBI, 2021).

Question 11. Outcome measures

Were the outcomes defined in detail? Were the tools or methods for measuring outcomes accurate and reliable–for example, have they been validated or are they objective? This issue is important because it influences confidence in the validity of study results. Also important is whether the outcomes were assessed in the same manner within groups and between groups.

An example of an outcome measure that is objective, accurate, and reliable is death–the outcome measured with more accuracy than any other. But even with a measure as objective as death, there can be differences in the accuracy and reliability of how death was assessed by the investigators. Did they base it on an autopsy report, death certificate, death registry, or report from a family member? Another example is a study of whether dietary fat intake is related to blood cholesterol level (cholesterol level being the outcome), and the cholesterol level is measured from fasting blood samples that are all sent to the same laboratory. These examples would get a "yes." An example of a "no" would be self-report by

subjects that they had a heart attack, or self-report of how much they weigh (if body weight is the outcome of interest).

Similar to the example in Question 9, results may be biased if one group (e.g., people with high BP) is seen more frequently than another group (people with normal BP) because more frequent encounters with the health care system increases the chances of outcomes being detected and documented (NHLBI, 2021).

Question 12. Blinding of outcome assessors

Blinding means that outcome assessors did not know whether the participant was exposed or unexposed. It is also sometimes called “masking”. The objective is to look for evidence in the article that the person(s) assessing the outcome(s) for the study (for example, examining medical records to determine the outcomes that occurred in the exposed and comparison groups) is masked to the exposure status of the participant. Sometimes the person measuring the exposure is the same person conducting the outcome assessment. In this case, the outcome assessor would most likely not be blinded to exposure status because they also took measurements of exposures. If so, make a note of that in the comments section.

As the reviewer assesses this criterion, he/she thinks about whether it is likely that the person(s) doing the outcome assessment would know (or be able to figure out) the exposure status of the study participants. If the answer is no, then blinding is adequate. An example of adequate blinding of the outcome assessors is to create a separate committee, whose members were not involved in the care of the patient and had no information about the study participants' exposure status. The committee would then be provided with copies of participants' medical records, which had been stripped of any potential exposure information or personally identifiable information. The committee would then review the records for prespecified outcomes according to the study protocol. If blinding was not possible, which is sometimes the case, mark "NA" and explain the potential for bias (NHLBI, 2021).

Question 13. Followup rate

Higher overall followup rates are always better than lower followup rates, even though higher rates are expected in shorter studies, whereas lower overall followup rates are often seen in studies of longer duration. Usually, an acceptable overall followup rate is considered 80 percent or more of participants whose exposures were measured at baseline. However, this is just a general guideline. For example, a 6-month cohort study examining the relationship between dietary sodium intake and BP level may have over 90 percent followup, but a 20-year cohort study examining effects of sodium intake on stroke may have only a 65 percent followup rate (NHLBI, 2021).

Question 14. Statistical analyses

Were key potential confounding variables measured and adjusted for, such as by statistical adjustment for baseline differences? Logistic regression or other regression methods are often used to account for the influence of variables not of interest.

This is a key issue in cohort studies, because statistical analyses need to control for potential confounders, in contrast to an RCT, where the randomization process controls for potential confounders. All key factors that may be associated both with the exposure of interest and the outcome–that are not of interest to the research question–should be controlled for in the analyses. For example, in a study of the relationship between cardiorespiratory fitness and CVD events (heart attacks and strokes), the study should control for age, BP, blood cholesterol, and body weight, because all of these factors are associated both with low fitness and with CVD events. Well-done cohort studies control for multiple potential confounders (NHLBI, 2021).

Discussion

The questions in the form of NIH-QATOCCsS are designed to help reviewers focus on the key concepts for evaluating the internal validity of a study. They are not intended to create a list that tally up to arrive at a summary judgment of quality.

Internal validity for cohort studies is the extent to which the results reported in the study can truly be attributed to the exposure being evaluated and not to flaws in the design or conduct of the study–in other words, the ability of the study to draw associative conclusions about the effects of the exposures being studied on outcomes. Any such flaws can increase the risk of bias.

Critical appraisal involves considering the risk of potential for selection bias, information bias, measurement bias, or confounding (the mixture of exposures that one cannot tease out from each other). Examples of confounding include co-interventions, differences at baseline in patient characteristics, and other issues throughout the questions above. A high risk of bias translates to a rating of poor quality. A low risk of bias translates to a rating of good quality (thus, the greater the risk of bias, the lower the quality rating of the study).

In addition, the more attention in the study design to issues that can help determine whether there is a causal relationship between the exposure and outcome, the higher the quality of the study. These include exposures occurring prior to outcomes, evaluation of a dose-response gradient, accuracy of measurement of both exposure and outcome, sufficient

timeframe to see an effect, and appropriate control for confounding–all concepts reflected in the tool.

Generally, when a reviewer evaluates a study, he/she will not see a "fatal flaw," but he/she will find some risk of bias. By focusing on the concepts underlying the questions in the quality assessment tool, the reviewer should ask himself about the potential for bias in the study that are critically appraising. For any box where reviewer check "no" he/she should ask, "What is the potential risk of bias resulting from this flaw in study design or execution?" That is, does this factor cause to doubt the results that are reported in the study or doubt the ability of the study to assess an association between exposure and outcome accurately?

The best approach is the reviewer to think about the questions in the tool and how each one tells something about the potential for bias in a study. The more the reviewer is familiarized with the key concepts, the more comfortable he/she will be with critical appraisal. Examples of studies rated good, fair, and poor are useful, but each study must be assessed on its own based on the details that are reported and consideration of the concepts for minimizing bias (NHLBI, 2021).

All these guidances are essential for determining the overall quality rating of observational cohort and cross-sectional studies.

Conclusions

The purpose of the present study was the official translation into Greek of the NIH-QATOCCsS proposed by the American National Institutes of Health. The tool is detailed, easy to use, with clear guidance, and its Greek translation is available for scientists and students writing in Greek, systematic reviews that include observational cohort or cross-sectional studies.

References

- Bagias, C., Sukumar, N., Weldeselassie, Y., Oyebode, O., & Saravanan, P. (2021). Cord blood adipocytokines and body composition in early childhood: a systematic review and meta-Analysis. *Int J Environ Res Public Health, 18*:1897. DOI: 10.3390/ijerph18041897
- Brislin, R.W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology 1*, 185–216.
- Brislin, R., W. (1976) Comparative research methodology: cross-cultural studies. *International Journal of Psychology 11*, 215–229.

- Drukker, M., Weltens, I., van Hooijdonk, C., F., M., Vandenberk, E., & Bak, M. (2021). Development of a methodological quality criteria list for observational studies: the observational study quality evaluation. *Front Res Metr Anal, 6*:675071. DOI: 10.3389/frma.2021.675071
- Jones, P., S., Lee, J.W., Pillips, L., R., Zhang, X., E., & Jaceldo, K., B. (2001). An adaptation of Brislin's translation model for cross-cultural research. *Nursing Research, 50*, 300–304. DOI: 10.1097/00006199-200109000-00008
- Sechrest, L., & Fay, T., L. (1972). Problems of translation in cross-cultural research. *Journal of Cross-Cultural Psychology 3*, 41–56.
- *United States government. National Heart, Lung, and Blood Institute (NHLBI)*. Study Quality Assessment Tools. The NIH - quality assessment tool for observational cohort and cross-sectional studies. (2021). Available at: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools