The Quantitative Cultural Validation of Three Tools among People Affected by Leprosy in West Java, Indonesia

In order to develop an internationally usable cross-NTD and cross-cultural toolkit of instruments to measure NTD-related morbidity and disability

Liesbeth Akkerman

Master student VU University
Student number: 2559517
Telephone: +316 222 99 108
Email: liesbeth_akkerman@hotmail.com
Supervisor: Dr. W. van Brakel
Master: Management, Policy Analysis and Entrepreneurship in the Health and Life Sciences
Specialization: International Public Health
Date: October, 2016, Amsterdam

EXECUTIVE SUMMARY

BACKGROUND

Neglected tropical diseases [NTDs] are a medical diverse group of 17 diseases that are strongly associated with poverty. Globally, an estimated 1 billion people are affected by NTDs. For example, in Indonesia most of the population live in areas where at least one NTD is endemic. The figures on morbidity and disability are only estimates, due to a lack of culturally valid tools. Previous research, done by Van 't Noordende et al. (2016), made a prototype toolkit with tools categorised according to the ICF domains. This study aims to quantitatively validate the P-scale Short, the WGQ and the SALSA. The tools relate to the ICF components *activity* and *participation*.

RESEARCH OBJECTIVE

The overall objective was to contribute to the development of an internationally usable cross-NTD and cross-cultural toolkit of instruments to measure NTD-related morbidity and disability, by a quantitative cultural validation study on persons affected by leprosy in West-Java, Indonesia.

THEORETICAL BACKGROUND

This study uses a framework bases on the work of Herdman et al. (1998), Stevelink & Van Brakel (2013) and Terwee et al (2007). The framework of this study consisted of five kinds of equivalences: four qualitative (conceptual, semantic, item, and operational) and one quantitative (measurement). Measurement equivalence refers to the psychometric properties of the tools. The validated psychometric properties in this study were construct validity, internal consistency, floor and ceiling effects, and interpretability.

METHODOLOGY

In order to validate scales quantitatively, first a qualitative check was needed. The qualitative check was done during the first three interviews. The quantitative validation was done with data from 101 interviews. The study population consisted of persons affected by leprosy living in a former leprosy community in Bantam, Java, Indonesia. The combination of the WGQ(6), WGQ(11), P-Scale Short and SALSA had several advantages: the tools are all based on the ICF, relate to the same or comparable ICF domains, are easy and quick to use, and have cross-culturally potential. Spearman's correlations, Cronbach's alphas, percentages of the lowest and highest scores, means, standard deviations, medians, and inter-quartile ranges were calculated in SPSS in order to the measurement equivalences. To prevent ethical inconvenience, ethical approval was obtained and all participants signed an informed consent.

RESULTS

The conceptual, semantic and item equivalence of the WGQ(6) and WGQ(11) was supported. During the pilot, item 4 was considered as confusing. After revising,

item 4 was together with the other items qualitatively valid. After all interviews, it was seen that most people needed explanation for question 7. Construct validity, internal consistency and floor and ceiling effects were for both WGQ versions appropriate, however the WGQ(11) scored better on all aspects. As example of interpretability is given how total mean score is significantly higher among participants with disability grade 1 or 2, compared with participants with disability grade 0, when measured with WGQ(11).

During the pilot, semantic; item; and operational equivalence was supported for the P-Scale Short. The measurement properties were considered as good. Question 12 was most difficult to understand, as most people needed an example or explanation.

For the SALSA scale, no problems were noticed regarding the conceptual, semantic or operational equivalence during the pilot. Question 2 led to confusion, and was therefore revised. The measurement properties were considered as good. Item 4 was most difficult to understand.

DISCUSSION

During the pilot, WGQ item 4 was revised. According to this and other studies, it is recommended to use the new version of the item. Semantic, item and operational equivalences are supported. When compared with the WGQ(6), the WGQ(11) is more comprehensive in facing all aspects of the construct activity, is more internal consistent, and is more sensitive in finding people facing activity limitations. Concluding, the WGQ(11) is more favourable to use in the study population.

The Bahasa Indonesia version of the P-Scale was already used in two large studies among persons affected by leprosy, which were partly conducted on Java. Therefore, remaining problems with the translation were unlikely. Indeed, this study did not find problems regarding semantic, item or operational equivalence. Although the internal consistency was lower compared with other studies, it still is considered as adequate. The other measurement equivalences were considered as good.

According to this and other studies, SALSA scale item 3 seemed to be less relevant when the answer option about risk avoidance is deleted. Therefore, this study recommends to remove question 3 if this answer option is removed as well. For question 2 is recommended to use the revised version and to make it two separate questions. The measurement properties of the SALSA scale are supported.

CONCLUSION

The overall conclusion is that this study shows that the WGQ(6), WGQ(11), P-Scale Short and SALSA are overall cultural valid to use among persons affected by leprosy in West-Java, while the WGQ(11) is in this study context more favourable than the WGQ(6). This study results are not one-to-one generalizable to other NTDs or regions: carefully testing or validating in the new context is needed. Persons who want to use these and other tools need to be highly aware of the importance of cultural validation of tools. More validation on a broader scale will be needed to develop an internationally usable cross-NTD toolkit of instruments to measure NTD-related morbidity and disability.

PREFACE

This internship was the type I desired to do: doing research in a foreign country for an NGO in the field of international public health. It was even better than I could wish on forehand. It was so interesting to live, work and make friends in a new culture. Unfortunately, my time in Indonesia was shorter than planned, as life cannot be planned as we might think sometimes. I would have loved to stay longer and hope to come back in Depok one day. Fortunately, my time in Indonesia was long enough to have gathered enough data for this report. I hope this research is a valuable next step in the development of an internationally useable NTD toolkit, as people affected by NTDs deserve attention for the challenging aspects they face in life. We cannot and may not leave people behind.

Without the people around me, both in Indonesia and The Netherlands, this internship would not have been possible. First of all, I would like to thank the Sitanala community; both community leaders with helping facilitating the interviews, and all participants who did their best to give answer on the personal questions of this research' tools. We felt so welcome in the community, which I am really thankful for. I wish you all the best! Second, many thanks to the Universitas Indonesia for helping us to find a place to live, an office, supervisors, interviewers and for making us feel home by, for example, the delicious daily lunches we had together at the office. Above all, thanks to Arafat for all your help. You helped me so much with your insights and always being there for me when I had a question again. Also special thanks to ibu Rita, pak Dadun, Gita and our friend Hendri.

Dr. Wim van Brakel: your quick, friendly and constructive way of giving feedback was so helpful. In addition, your expertise, perspectives and good work in the NTD field is an inspirational example for my own future career. Your were the best supervisor I could wish for. Thank you!

Roline, it was so much fun to share life with you in Depok! I could talk with you about challenges faced during doing research, but above all you became a dear friend to me during all fun, personal talks and activities we have done together.

So many thanks to my dear parents for making it possible to study and to go abroad, which helped me to grow in both professional and personal domains. With your everlasting love, you encouraged and strengthened me to make the best out of myself. And you Bart-Jan, for loving me so much and encouraging me in my (abroad) ambitions; your unconditional love and support makes it possible for me to live my dream, as I did with this internship.

Liesbeth Akkerman

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CHIEF: Craig Hospital Inventory of Environmental Factors
CI: confidence interval
CRPD: Convention on the Rights of Person with Disabilities
DALY: Disability-Adjusted Life Years
EMIC: Explanatory Model Interview Catalogue
IQR: inter-quartile range
NLR: Netherlands Leprosy Relief
NMD: NTD related morbidity and disability
NTD: neglected tropical disease
P-scale short: shortened version Participation scale
SALSA: Screening of Activity Limitation and Safety Awareness
SD: standard deviations
SRQ: Self-Reporting Questionnaire
UI: Universitas Indonesia
UN: United Nations
VU University: Vrije Universiteit
WGQ: Washington Group short set of questions
WGQ(6): Washington Group short set: 6 core questions
WGQ(11): Washington Group set of questions: 6 core + 5 additional questions
WHO: World Health Organisation
WHODAS: WHO Disability Assessment Schedule
WHOQOL-BREF: WHO Quality of Life (abbreviated)
WHOQOL-DIS: WHO Quality of Life (disability)

1   INTRODUCTION

Neglected tropical diseases [NTDs] are a group of 17 diseases that are medically diverse, but form a group as all are strongly associated with poverty, all flourish in impoverished environments and all thrive in tropical areas (WHO, 2010). The term "neglected" is used in the name of this group diseases, as the diseases received little or no public health attention, advocacy or funding, until recently (Kolaczinski, Kabatereine, & Onapa, 2007). Although several NTD-related programmes exist nowadays, still enormous progress is needed (UN, 2006; WHO, 2001, 2011).

NTDs exist in 149 countries in southeast Asia, east Asia, sub-Saharan Africa and Latin America, where at least 100 countries are endemic for 2 of more NTDs (WHO, 2010). For example, in Indonesia, soil-transmitted helminth infections, lymphatic filariasis, schistosomiasis, dengue, yaws and leprosy are co-endemic in at least parts of the country (Tan, Kusriastuti, Savioli, & Hotez, 2014). Approximately 195 million Indonesian citizens, which is most of the population, live in areas where at least one NTD is endemic (Tan et al., 2014).

Estimates are that NTDs cost the lives of around 500,000 people per year (Hotez, Ottesen, Fenwick, & Molyneux, 2006). Although these numbers are already striking, the main impact of NTDs is not captured by mortality figures, but arises from chronic disability and morbidity[1] (Hotez, Ottesen, et al., 2006). Globally, an estimated 1 billion people are affected by NTDs, causing a global burden of 56.6 million DALYs[2] (WHO, 2010, 2015). It is estimated that only lower respiratory infections, HIV/AIDS, and diarrheal diseases cause more DALYs globally (Hotez, Molyneux, Fenwick, & Ottesen, 2006).

The figures on morbidity and disability are only estimates, as measurements are complex and little tools are available. One reason is that morbidity and disability result from an interaction of different aspects, namely health condition, environment and personal factors (Madans & Loeb, 2011; WHO, 2001). Therefore, NTD morbidity and disability [NMD] data may relate to different aspects of life and may be collected for different goals, which makes it difficult to compare data (Madans & Loeb, 2011). Comparable data is needed for fundraising, needs assessments, intervention planning, programme management, and monitoring and evaluation of interventions. Therefore, there is consensus about the need for standard methods to measure NMD.

---

[1] Morbidity, definition: "Suffering due to a disease or other conditions." (WHO, 2013, p. viii)

[2] DALY, definition: "One DALY can be thought of as one lost year of "healthy" life. The sum of these DALYs across the population, or the burden of disease, can be thought of as a measurement of the gap between current health status and an ideal health situation where the entire population lives to an advanced age, free of disease and disability." (WHO, 2014, p. 1)

The first steps towards a consensus on methods to assess and monitor NMD were made by conventions, organisations and researchers. An important convention was one of the United Nations [UN] in 2006: the Convention on the Rights of Person with Disabilities [CRPD], aiming to improve the situation of disabled persons in several ways, such as monitoring their situation (UN, 2006). An organisation that helps to improve measurement of NMD is the Netherlands Leprosy Relief [NLR]. Researchers Van 't Noordende and colleagues (2016) aimed to develop a cross-cultural toolkit of instruments to assess and monitor NMD. This toolkit is the foundation for this study.

The toolkit consists of eight tools that were selected and recommended by professionals and experts in NTDs and disability. Tools that are not yet included in the toolkit but might be valuable to include are the Washington Group Questionnaire [WGQ] and the Screening of Activity Limitation and Safety Awareness [SALSA] scale. The Washington Group aimed to define and measure disability in a way that is cross-cultural applicable and appropriate in several contexts (Madans & Loeb, 2011). To achieve this, they made a short set of six disability-related questions: the WGQ (Madans & Loeb, 2011). Furthermore, the Salsa Collaborative Study Group aimed to develop a method of measuring activity limitation in diseases such as leprosy (SALSA Collaborative Study Group, 2007). This resulted in the SALSA scale.

Before the NMD toolkit of Van 't Noordende et al. (2016), the WGQ and the SALSA scale can be used in an international and cross-NTD manner, the tools have to be validated in several cultures, languages, and NTDs, both qualitatively and quantitatively (Bowden & Fox-Rushby, 2003). This study aims to quantitatively validate the Participation Scale Short [P-scale Short], which is part of the prototype NMD toolkit, the WGQ and the SALSA scale among persons affected by leprosy on West-Java, Indonesia. This study is part of a larger study, in which also other tools of the prototype NMD toolkit are validated among several NTDs and in several countries.

OVERALL OBJECTIVE
To contribute to the development of an internationally usable cross-NTD and cross-cultural toolkit of instruments to measure NTD-related morbidity and disability by a quantitative cultural validation study on persons affected by leprosy in West-Java, Indonesia.

2   CONTEXTUAL BACKGROUND

2.1 ICF

The International Classification of Functioning, Disability and Health [ICF] (figure 1) is a framework made by the World Health Organisation [WHO] (WHO, 2001). The ICF consists of the following components: health condition (disease or disorder); body functions and structure; activity; participation; contextual factors (environmental or personal). The ICF shows the interactive relation between the different components, as all components are related to each other in all directions. To map the health condition of a person, all different domains should be taken into account. As every component has own specific features, every component has to be measured specifically.

The tools that were culturally validated in this study relate to the components *activity* and *participation*. Activity "is the execution of a task or action by an individual" (WHO, 2001, p. 14). Participation "is involvement in a life situation" (WHO, 2001, p. 14). Both activity and participation cover the level of how much a person is able to perform actions: functioning from both an individual and a societal perspective, respectively (Kostanjsek, 2011). As activity and participation may relate to the same life domains, tools that measure activity may also give information about aspects of participation, and -vice versa- participation about activity.

      Tools can assess activity limitations and participation restrictions of a person. Activity limitations are "difficulties an individual may have in executing activities" (WHO, 2001, p. 14). Participation restrictions are "problems an individual may experience in involvement in life situations" (WHO, 2001, p. 14). The measurement outcomes can be compared to those of a person without a similar health condition (WHO, 2001). The difference in outcomes of someone who is affected by a disease, for example leprosy, and someone who is not affected by a disease may show scope for improvement.
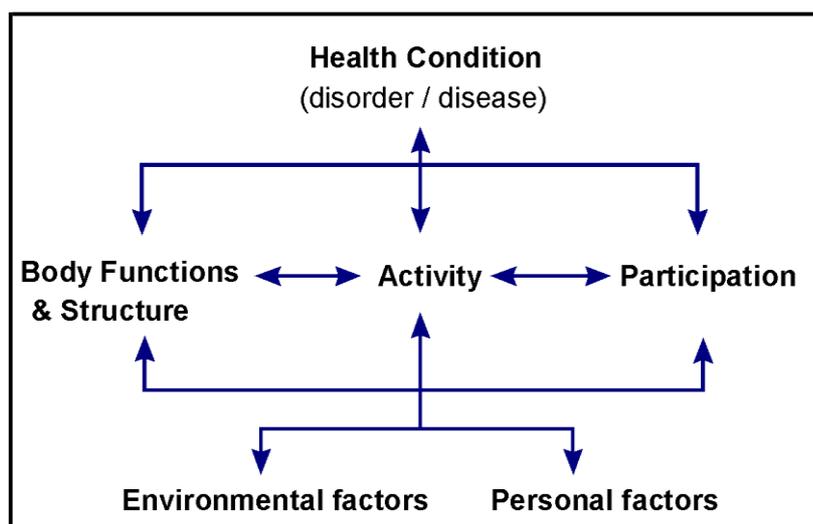
Figure 1. International Classification of Functioning, Disability and Health [ICF]

2.2 PREVIOUS RESEARCH NMD TOOLKIT

In 2015, a prototype NTD-related morbidity and disability [NMD] toolkit was developed through a literature review of existing tools, followed by a Delphi study with NTD experts (Van 't Noordende et al., 2016) (Figure 2). The tools were categorised according to the ICF domains in such a way that every domain was covered. For example, the Participation Scale (P-scale) covered the ICF domain *participation*.

The first validation of the toolkit showed promising results, as five out of six validated tools were potentially suitable for assessing and monitoring different aspects of NMD in Northeast Brazil (Van 't Noordende et al., 2016). However, to develop a toolkit that is useful cross-culturally and across NTDs, the tools of the NMD toolkit should be validated in various countries and among persons affected by various NTDs (Bowden & Fox-Rushby, 2003).

ICF Domain	Tool
Impairments (body functions)	Clinical profile
Participation restrictions	P-scale (short)
Activity limitations	WHODAS 2.0
Impairments; mental health problems	SRQ
Personal factors; quality of life	WHOQOL-BREF and DIS
Environmental factors	CHIEF
Environmental factors and personal factors; stigma	EMIC

CHIEF: Craig Hospital Inventory of Environmental Factors; EMIC: Explanatory Model Interview Catalogue; P-scale: Participation scale; SRQ: Self Reporting Questionnaire; WHODAS: WHO Disability Assessment Schedule; WHOQOL-BREF: WHO Quality of Life (abbreviated); WHOQOL-DIS: WHO Quality of Life (disability).

Figure 2. Prototype NMD toolkit (Van 't Noordende et al., 2016)

3 THEORETICAL BACKGROUND

This section shows and explains the framework used for this study and gives the research questions.

3.1 FRAMEWORK

The research objective of this study is to quantitatively and culturally validate instruments to measure NTD-related morbidity and disability on persons affected by leprosy in West-Java, Indonesia. Herdman, Fox-Rushby, & Badia (1998) provided a basis for critiquing and reviewing methodologies used for cross-cultural research. This study uses a framework that draws mainly on the work of Herdman et al. (1998), as Peters et al. (2014) and Stevelink & Van Brakel (2013) also did in their studies. Stevelink & Brakel (2013) provided a checklist to test the cultural validity. This checklist is used in this study. The first part of this checklist concerns qualitative and the second part quantitative validation. The quantitative part includes psychometric properties, which are based mainly on Terwee et al. (2007). The study of Terwee et al. (2007) is useful for this study, as they provided explicit quality criteria for outcomes of studies that aim to quantitatively validate health related tools.

3.1.01 QUALITATIVE ASPECTS

This study evaluated four kinds of qualitative equivalences of tools: conceptual, semantic, item, and operational equivalence. The types of equivalences are explained in the following section, including the way we validated the equivalences. The corresponding interview guide can be found in 8.2.02.

CONCEPTUAL EQUIVALENCE

Investigating conceptual equivalence essentially involves exploring the ways in which different populations conceptualize particular concepts (Herdman et al., 1998). According to the universalist approach, the relevance of concepts may differ between cultures (Herdman et al., 1998). In a (sub-)culture where the tool has not yet been used, concepts should be qualitatively validated before the instruments can be quantitatively validated on a larger scale. Only translating an English tool to another language, such as Bahasa Indonesia in this study, is not enough. Conceptual equivalence of the tools is achieved when the construct employed in the original questionnaire is equally valid in the target culture (Herdman et al., 1998).

During preparations of this study, literature search showed that the tools are already used in studies in Indonesia. From this and other studies we concluded that there were no reasons to believe that the concepts *activity* and *participation* in the study population were different than in the original culture. Therefore, no further research was done regarding conceptual equivalence.

SEMANTIC EQUIVALENCE

According to Herdman et al. (1998, p. 326), "semantic equivalence is concerned with the transfer of meaning across languages, and with achieving a similar effect on respondents in different languages". Semantic equivalence takes into account that a word may have different types of meaning, between and within cultures. An important aspect of semantic equivalence is that the target population understands the language used well and that the original meaning behind a word, item or question has been preserved.

We validated the semantic equivalence with help of both professionals and participants. Professionals who speak both English and Bahasa language and who understand tools' concepts and items looked at the Bahasa versions of the tools. They studied if the items in the Bahasa tools had the same meaning as the English version, if the words and questions were acceptable to use, and if the language was not too difficult to understand for participants with low literacy. After this check, the first three participants were also asked if they found the questions easy to understand, and if they found all words acceptable.

ITEM EQUIVALENCE

Item equivalence asks whether the relevance and acceptability of items is similar between cultures (Herdman et al., 1998). "Items in the questionnaires must reflect areas that are important to the target population" (Terwee et al., 2007, p. 35). Items may in fact ask very different things in different cultures (Herdman et al., 1998). For example, being less able to work may indicate very different levels of incapacity and/or emotional impact in different cultures. Also, items can vary in acceptability (Herdman et al., 1998). Items that are acceptable to ask in one culture, might be rude to ask or a taboo in another culture.

We validated the item equivalence of the tools by asking the first three participants about relevance an acceptability of the items.

OPERATIONAL EQUIVALENCE

Operational equivalence concerns the suitability of the questionnaire format, instructions and mode of administration (Peters et al., 2014). It takes into account whether the same questionnaire format, instructions and mode of administration can be used in the target population without affecting the responses (Herdman et al., 1998). Operational aspects that might influence responses are for example: the literacy rate for written questionnaires, familiarity with response scales (e.g. the Likert scale), and time frames (e.g. a person who is asked to think of the previous week) (Herdman et al., 1998). All three tools in this study use a version of the Likert scale.

We validated the operational equivalence by observing if participants were able to answer the tool questions in the way they were asked and whether they understood the response options easily. By doing this, the item was marked if the participant asked for an example or explanation for it. After asking the questions, the first three participants were asked what they think of the way the questions were asked and the answer options. If they showed difficulties with answering questions, suggestions were asked how the questions could have been better formulated. Furthermore, these first three participants were asked if the duration of the interview was acceptable.

3.1.02 QUANTITATIVE ASPECTS

Measurement equivalence refers to the psychometric properties of the tools. According to Terwee et al. (2007), these are: construct validity, internal consistency, reproducibility, floor and ceiling effects, and interpretability. Due to time constrains, measuring reproducibility was not possible in this study.

CONSTRUCT VALIDITY

Construct validity "refers to the extent to which scores on a particular instrument relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured" (Kirshner & Guyatt, 1985). As the WGQ and the SALSA scale both measure activity, the hypothesis was that scores were closely correlated. Also P-scale scores in relation to WGQ and SALSA scores were expected to show correlation, but this relationship was not to be one-to-one (Velema, 2010).

INTERNAL CONSISTENCY

Internal consistency is a measure of the extent to which items in a questionnaire scale are correlated (Terwee et al., 2007). An instrument is internal consistent when it consists of multiple items measuring the same concept (Terwee et al., 2007).

FLOOR AND CEILING EFFECTS

According to Terwee et al. (2007), floor and ceiling effects were considered to be present if more than 15% of respondents achieved the lowest or highest possible score, respectively. If floor or ceiling effects are present, it is likely that extreme items are missing in the lower or upper end of the scale, indicating limited content validity (Terwee et al., 2007). As a consequence, participants with low or high scores cannot be distinguished from each other, which reduces reliability (Terwee et al., 2007).

INTERPRETABILITY

According to Lohr, Aaronson, Alonso, & Burnam (1996, p. 981), interpretability is defined as "the degree to which one can assign qualitative meaning to

quantitative scores". Researchers who aim to develop or validate measurement tools should publish results that help to assess how the scores should be interpreted (Terwee et al., 2007). Examples include comparative data regarding distribution, means and standard deviations of scores among different groups. Groups could be compared according to age, educational level, employment, or disease conditions, for example.

Conceptual equivalence	Exploring the ways in which different populations conceptualize concepts
Semantic equivalence	Translation well understood; achieving a similar meaning and effect on respondents in different languages
Item equivalence	Differences of acceptability and relevance of items between cultures.
Operational equivalence	Possibility of using the same operational formats, instructions and mode of administration.
Measurement equivalence	Refers to the psychometric properties: ○ Construct validity ○ Internal consistency ○ Floor and ceiling effects ○ Interpretability

Figure 3. Framework

3.2 RESEARCH QUESTIONS

The objective of this study is:
- To contribute to the development of an internationally usable cross-NTD toolkit of instruments to measure NTD-related morbidity and disability *by* conducting a quantitative cultural validation study among persons affected by leprosy in West-Java, Indonesia.

The overall research question is:
- How cultural valid are the two version of the WGQ, the P-Scale Short and the SALSA scale among persons affected by leprosy in West-Java, Indonesia?

The sub-questions are:
- To what extent are the two versions of the WGQ equivalent to the original version in terms of semantic, item, operational and measurement equivalence when used among persons affected by leprosy in West-Java, Indonesia?
- To what extent is the P-Scale Short equivalent to the original version in terms of semantic, item, operational and measurement equivalence when used among persons affected by leprosy in West-Java, Indonesia?

- To what extent is the SALSA scale equivalent to the original version in terms of semantic, item, operational and measurement equivalence when used among persons affected by leprosy in West-Java, Indonesia?

4 METHODS

Section 4.1 describes the study design; section 4.2 the study population and study sample; section 4.3 the sample size and sampling method; section 4.4 the data collection; section 4.5 the outcome measures; section 4.6 the data analysis and management; section 4.7 the resources; and section 4.8 the ethical considerations.

4.1 STUDY DESIGN

In order to validate scales quantitatively, first a qualitative check was needed.

4.1.01 QUALITATIVE RESEARCH

The interview guide for this qualitative part can be found in section 8.2.02. We used an interview guide including topics that were based on the previous NTD-toolkit validation study in Brazil (Van 't Noordende et al., 2016) and also based on the first part of the framework made by Herdman et al. (1998): conceptual, item, semantic and operational equivalence.

4.1.02 QUANTITATIVE RESEARCH

The quantitative research that was done was about the measurement equivalence, which in this study included the following psychometric properties: construct validity, internal consistency, floor and ceiling effects and interpretability. The quantitative part was based on the study of Herdman et al. (1998), Stevelink & Van Brakel (2013) and Terwee et al. (2007).

4.2 STUDY POPULATION AND STUDY SAMPLE

The study population consisted of persons affected by leprosy living in West-Java, Indonesia.

4.2.01 INCLUSION CRITERIA

Participants were eligible if they:
- o Have been diagnosed with leprosy
- o Were 16 years or above
- o Spoke Bahasa Indonesia with sufficient fluency
- o Gave written or verbal informed consent
- o Were an Indonesian citizen from Java

4.2.02 EXCLUSION CRITERIA

Participants were not eligible if they:
- o Have been diagnosed with other disabling and/or stigmatized diseases as well.

4.3 SAMPLE SIZE AND SAMPLING METHOD

The qualitative check was done during the first three interviews with persons affected by leprosy. The quantitative validation was done with 100 persons affected by leprosy (Terwee et al., 2007). Every participant answered questions of all tools. In addition, they answered questions about their demographic features (see 8.2.01).

To find enough participants who were willing to participate, contact was made with a former leprosy community Sitanala Community (Bantam, Java, Indonesia). As this was a former leprosy community with still a high-density of people affected by leprosy, this study population differed from the general community in key aspects, such as disability. The possibility that the target community was not representative for the general population on Java is no problem, as the goal of the study was to validate the questionnaires, not to measure the activity/participation restriction on Java.

4.4 DATA COLLECTION

4.4.01 INTERVIEW GUIDES

The interview guide guided the collection of qualitative data (8.2). The quantitative data was collected by asking participants the instruments' questions. The instruments were used in a structured way, so it led to comparable data. During the interviews, the interviewers made notes about duration of the interviews, if people needed an example and/or explanation or anything else noticeable.

4.4.02 TOOLS

In section 8.1, the possibly relevant tools can be found. This list includes the prototype NMD toolkit of Van 't Noordende et al. (2016), the short set of questions made by the Washington Group [WGQ] (Madans & Loeb, 2011), and the SALSA scale (SALSA Collaborative Study Group, 2007). The tools were screened with regard to earlier validation or use in Indonesia among persons affected by NTDs, whether validation was needed among persons affected by leprosy, and whether there was a Bahasa version of the tool available. Some of the tools that included in this list are, and some are not used in Indonesia yet. As this study focuses on quantitative validation, it was necessary that the tools had already been qualitatively validated and/or used without problems in the Indonesian context.

The tools that were used in this study were two versions of the WGQ (section 8.4), P-scale Short (section 0), and SALSA (section 8.6). Some tools were excluded, as there was no Bahasa Indonesia version available (Clinical Profile, CHIEF,

WHOQOL-DIS). It could be valuable to validate the Bahasa versions of the other tools also, such as the SRQ, WHODAS 2.0, WHOQOL-BREF and EMIC-CSS. However, due to time constraints, no more than three tools could be included.

The combination of the three tools had several advantages. The three tools are all based on the ICF and relate to the same or comparable ICF domains: participation and activity. Furthermore, all three tools are made to be easy and quick to use, and are developed to be used cross-culturally.

WGQ

The Washington Group is a UN voluntary working group made up of representatives from governments, non-governmental organisations and disability organisations. The Washington Groups aims "to deal with the challenge of disability definition and measurement in a way that is culturally neutral and reasonably standardized among the UN member states" (Madans & Loeb, 2011, p. 1). To achieve this, they made a short set of six questions to measure disability, which we call WGQ(6). Additional questions on the six core questions are recommended to get a more comprehensive overview of participants' activity limitation (Madans & Loeb, 2011). In Bahasa Indonesia, five additional questions existed as extension of the six core WGQ questions (Van Brakel et al., 2012). These five questions are together with the six core questions the county-specific WGQ we validated in Indonesia (see section Table 11). We called this extended version the WGQ(11).

Each WGQ question has four response categories: no – no difficulty; yes – some difficulty; yes – a lot of difficulty; and cannot do at all. The responses scales are scored from 0-4, where "no – no difficulty" is scored as 0 and "cannot do at all" as 4. With these response categories, respondents are able to capture the level of activity limitation from mild to severe (Madans & Loeb, 2011).

P-SCALE SHORT

Van Brakel & Anderson (2006) designed the P-scale as they discovered a lack of tools to measure participation. Stevelink, Hoekstra, & Nardi (2012) designed a shortened version of the P-scale. This P-Scale Short was already included in the prototype NMD-toolkit (Figure 2).

The P-scale Short is a 13-item interview-based instrument. Stevelink et al. (2012) describes how the scale is used. The scale has a two-tier question and response format. First, the respondent is asked whether he or she experiences restriction in a particular aspect of participation. Some questions ask respondents to compare themselves with a peer. The possible answers are: yes, sometimes, no and irrelevant/I don't want to/don't have to. If the respondents answer "sometimes" or "no", the respondent is asked how big a problem the restriction is to him or her. The possible answers options are: no problem (1 point), small problem (2

points), medium problem (3 points) and large problem (5 points). All answers are scored as 0-5. The overall score is the sum of the individual item scores. A higher score indicates a higher level of participation restriction.

SALSA

The short questionnaire for screening of activity limitation and safety awareness [SALSA] is developed to measure activity limitation in diseases characterized by peripheral neuropathy, such as leprosy and diabetes, in a quick and simple way (SALSA Collaborative Study Group, 2007). The questionnaire consists of 20 questions.

The SALSA Collaborative Study Group provided an users' manual on how to use the SALSA scale and also a question-by-question guide (Velema, 2010). Every question has the response options "yes" or "no". Next, a follow-up question has to be asked. If the respondent answers "yes", the follow-up question "how easy is it for you?" with answer options: "easy"; "a little difficult" or "very difficult". If the respondent answers "no", the follow-up question is "why not?" with the answer options: "I don't need to do this", "I physically cannot", or "I avoid because of risk".

4.5 OUTCOME MEASURES

4.5.01 QUALITATIVE PART

CONCEPTUAL EQUIVALENCE

Conceptual equivalence was confirmed, as there were no reasons to doubt the similarity of the concepts in the Javanese culture compared to the original culture.

SEMANTIC EQUIVALENCE

The outcome of measuring semantic equivalence was whether professionals and the first three participants found the words acceptable to use and easy to understand.

ITEM EQUIVALENCE

Item equivalence was confirmed, if during the pilot study there was no inconvenience about the acceptability and relevance of the items. In addition to this qualitative way of measuring item equivalence, it was also measured in a quantitative way. The quantitative way used the interviewers' notes about if respondents needed explanation or an example.

OPERATIONAL EQUIVALENCE

The measurement outcome of operational equivalence was the opinion of the participants about the ability to answer the tool questions, whether they understood the response options easily, and about the duration of the interview.

4.5.02 QUANTITATIVE PART

CONSTRUCT VALIDITY

Measurement of the construct validity was done with Spearman's correlation. The strength of the correlation can be verbally described with using the following guide for the absolute value of r_s: .00-.19 = very weak; 0.20-0.39 = weak; 0.40-0.59 = moderate; 0.60-0.79 = strong; 0.80-1.0 = very strong. The predefined hypothesis to assess the construct validity was a moderate correlation between the WGQ(6) – SALSA, WGQ(11) – SALSA and WGQ(11) – P-Scale Short, WGQ(6) – P-Scale Short, and P-Scale Short - SALSA (Peters et al., 2014; Velema, 2010).

INTERNAL CONSISTENCY

Measurement of the internal consistency was done with Cronbach's alpha. An $0.9 \leq \alpha < 0.95$ was classified as excellent, $0.7 \leq \alpha < 0.9$ as good, $0.6 \leq \alpha < 0.7$ as acceptable, $0.5 \leq \alpha < 0.6$ as poor and $\alpha < 0.5$ as unacceptable.

FLOOR AND CEILING EFFECTS

Floor and ceiling effects were measured by assessing the percentage of participants who scored the lowest or highest score, respectively (Terwee et al., 2007). There were floor or ceiling effects if more than 15% of participants achieved the lowest or highest score, respectively (Terwee et al., 2007).

INTERPRETABILITY

To promote the interpretability of the scores in our study, we calculated the means and standard deviations [SD] of scores of all participants, scores of different sub-groups according to age, educational level, employment and disability grade (Terwee et al., 2007).

ITEMS

It was assumed that the data did not follow a normal distribution. Therefore, median values and inter-quartile range (IQR) were calculated. The IQR indicates the extent to which the central 50% of values within the dataset (the upper quartile minus the lower quartile) are dispersed. The IQR is based upon, and related to, the median, which is the value separating the higher half of the dataset from the lower half. The median and IQR were calculated per item to show how the scores of each question were distributed.

4.6 DATA MANAGEMENT AND ANALYSIS

During the pilot interviews, notes were made of participants' answers on the semi-structured pilot questions. Conversations with professionals and input of the interviewers also let to insight into the qualitative equivalences.

The quantitative part is analysed in SPSS. Several statistical tests are used to measure the instrument properties. See 4.5 for more details.

All data is stored on two laptops and an external hard drive in locked files.

4.7 RESOURCES

4.7.01 PERSONNEL

- o 1 Student, Master program, VU University Amsterdam: Ms. L. Akkerman. Worked 6 months full time on this project; 4 months in The Netherlands and 2 in Indonesia, Java.
- o 1 Student, Postgraduate Program, Universitas Indonesia: Mr. A. Patria. Supported the project by leading the interviews, by interpreting in various situations, by sharing knowledge and experience and by being involved in the analysis.
- o 1 Supervisor, Universitas Indonesia, Mr. Dadun. Supervised the on-site work and data collection.
- o 1 Supervisor, Netherlands Leprosy Relief, Dr. W. H. van Brakel. Supervised the preparation of the protocol and report writing; advised and gave feedback from a scientific and experience perspective.

4.7.02 FINANCES

Both researchers received a financial grant of €500 from the NLR. Besides this grant, the Dutch researcher paid all costs herself (institution fees; flight tickets; vaccinations; accommodation; visa, et cetera).

4.8 ETHICAL CONSIDERATIONS

We did not expect that participants in our study were harmed in any way. We did everything we could to avoid and minimalize any inconvenience to the participants. Before every interview, the participant was informed about the study content and goal, and about the rights he/she has as a participant. For example, participants have the right to not answer a question, without giving a reason. See 8.1 for more details. When participants indicated they understood the study content and goal and of their rights as a participant, informed verbal consent was asked.

To ensure that the study followed ethical procedures at an acceptable and scientific level, ethical approval was sought and obtained from the ethics committee of the UI public health faculty.

5 RESULTS

In this section, the results are described. Section 5.1 gives the characteristics of the study population. Section 5.2 gives the results regarding the qualitative equivalences per tool. Section 5.3 gives the results regarding the quantitative equivalences per tool.

5.1 CHARACTERISTICS STUDY POPULATION

All interviews are conducted at the Sitanala Community (Bantam, Java, Indonesia), which is a former leprosy community where still a lot of people are affected by leprosy. This study included 114 interviews. Of these, 13 interviews were omitted due to problems with one interviewer (10) and the first interviews were used as a pilot (3). The remaining 101 interviews were included in this study.

The socio-demographic features of the respondents are shown in Table 1. More men (57%) participated than women (43%). All participants were affected by leprosy; 43% had a grade 0 disability, 22% grade 1, and 36% a grade 2 disability[3]. The main age was 50 (SD: 12, range: 29-99). Most participants had at least some education (89%). One third (32%) did not finish primary school. Nobody had higher education than finishing senior high school. Most participants were married (83%). Most participants worked (63,2%), mostly as a self-employed worker (25.7%).

Variables		Respondents (n=101)
Sex	Men	58 (57.4%)
	Women	43 (42.6%)
Disability grade	Grade 0	43 (42.6%)
	Grade 1	22 (21.8%)
	Grade 2	36 (35.6%)
Age	Mean (SD; range)	50 (SD: 12; range: 29-99)
	< 31 years	4 (4.0%)
	31-45 years	36 (35.6%)
	46-60 years	44 (43.6%)
	61-75 years	13 (12.9%)
	> 75 years	4 (4.0%)
Education	No education	11 (10.9%)
	Stopped during primary school	32 (31.7%)
	Finished primary school	28 (27.7%)
	Stopped during junior high school	1 (1.0%)
	Finished junior high school	14 (13.9%)
	Stopped during senior high school	2 (2.0%)
	Finished senior high school	13 (12.9%)
Marital status	Single	7 (6,9%)
	Married	84 (83.2%)
	Divorced	4 (4.0%)

[3] Grade 0 = no impairment found; grade 1 = loss of sensation in the hand or foot; grade 2 = visible damage is noted (SALSA Collaborative Study Group, 2007).

Variables		Respondents (n=101)
	Widowed	7 (6.9%)
Current job	No work, because of health reasons	9 (8.9%)
	No work, because of other reasons	5 (5.0%)
	Stay at home mom	21 (20.8%)
	Self-employed	26 (25.7%)
	Government/Salaried employed	17 (16.8%)
	Daily labour/Hourly employee	21 (20.8%)
	Volunteer	2 (2.0%)

Table 1. Demographic features participants

5.2 QUALITATIVE PART

5.2.01 WASHINGTON SET OF QUESTIONS

The WGQ(6) and WGQ(11) questionnaires can be found in section 8.4.

CONCEPTUAL EQUIVALENCE

There were no reasons to doubt the existence of a concept of *activity* in the study population different than in the original culture.

SEMANTIC EQUIVALENCE

Although the participants often had a different language than Bahasa Indonesia as their mother tongue, everybody was able to understand the Bahasa Indonesia used in the interviews.

ITEM EQUIVALENCE

The Bahasa Indonesia version asked in question 4 both the domains communication and cognition. The original Bahasa question was:

> Originally included question Bahasa version:
>
> - Do you have troubles remembering or communicating with others due to your physical or mental condition?[4]

It might be confusing to answer a question that relies on two different topics: both cognition and communication. Therefore, with the help of a local professional, we made two questions out of this question. By doing this, the original WGQ was kept in mind, as this one has also two separate questions for the two domains.

> Questions original English version WGQ:

---

[4] Bahasa Indonesia: Apakah anda mengalami kesulitan mengingat atau berkonsentrasi atau berkomunikasi dengan orang lain karena kondisi fisik atau mental?

> - Do you have difficulty remembering or concentrating?
> - Using your usual (customary) language, do you have difficulty communicating, for example understanding or being understood?

The two new Bahasa questions became:

> Newly included questions of Bahasa version:
>
> - Do you have difficulty understanding or being understood when communicating, using daily language?[5]
> - Do you have difficulty remembering or concentrating?[6]

These two newly translated questions did not lead to any problems during the pilot interviews and therefore replaced the question that was originally in the Bahasa version.

During the pilot study and training of interviewers, no problems popped up regarding other items.

OPERATIONAL EQUIVALENCE

Participants were able to answer the tool questions, as they understood the operational format, instructions and mode of administration easily. The duration of conducting the WGQ(11) was on average 3.4 minutes (mode: 3 minutes, range: 2-7 minutes). The duration of the WGQ(6) is not measured.

5.2.02 P-SCALE SHORT

CONCEPTUAL EQUIVALENCE

There were no reasons to doubt the existence of a concept of *participation* in the study population different than in the original culture.

SEMANTIC EQUIVALENCE

Although the participants often had a different language than Bahasa Indonesia as their mother tongue, everybody was able to understand the Bahasa Indonesia used in the interviews.

ITEM EQUIVALENCE

During the pilot study and training of interviewers, no problems popped up regarding the items.

---

[5] Bahasa Indonesia: Apakah anda mengalami kesulitan memahami dan atau dipahami pada saat berkomunikasi dengan menggunakan bahasa sehari-hari?

[6] Bahasa Indonesia: Apakah anda mengalami kesulitan mengingat atau berkonsentrasi?

Before conducting the P-Scale Short, the interviewer explained the peer concept to the respondent carefully. Once the respondents could think of a peer in their life, the respondents were able to answer the tool questions, as they understood the operational format, instructions and mode of administration easily. The duration of the P-Scale Short was on average 6.8 minutes (mode: 5 minutes, range: 4–15 minutes).

5.2.03 SALSA

CONCEPTUAL EQUIVALENCE

There were no reasons to doubt the existence of a concept of *activity* in the study population different than in the original culture.

SEMANTIC EQUIVALENCE

Although the participants often had a different language than Bahasa Indonesia as their mother tongue, everybody was able to understand the Bahasa Indonesia used in the interviews.

ITEM EQUIVALENCE

During the training of the interviewers, the comment was made that participants may find it difficult to answer Question 2 of the SALSA:

Originally included question of Bahasa version:

- Do you sit or squat on the ground?[7]

The interviewers thought participants might get confused and see it as a redundant question as the interviewer can see whether the participant is sitting while (s)he asked this question. To overcome this problem, we changed "do" into "can". Question 2 of the SALSA became:

Newly included questions of Bahasa version:

- Can you sit or squat on the ground?[8]

The question about sitting or squatting on the ground[9] led still sometimes to confusion, as participants might understand sitting differently from squatting on the ground, whereby to squat on the ground was seen as more difficult than sitting.

---

[7] Bahasa Indonesia: Apakah anda jongkok dan duduk di lantai?

[8] Bahasa Indonesia: Dapatkah anda jongkok dan duduk di lantai?

[9] Bahasa Indonesia: Dapatkah anda jongkok dan duduk di lantai? / English: Can you sit or squat on the ground?

Participants were able to answer the tool questions, as they understood the operational format, instructions and mode of administration easily. The duration of the SALSA was on average 5.9 minutes (mode: 5 minutes, range: 3-13 minutes).

5.3 QUANTITATIVE PART
5.3.01 WASHINGTON SET OF QUESTIONS

Corresponding table of this section is Table 2.

CONSTRUCT VALIDITY
We found a strong correlation between the WGQ(11) and the SALSA (rho = 0.64) and a moderate correlation between the WGQ(11) and the P-Scale Short (rho = 0.42). Both correlations are significant at the 0.01 level.

We found a moderate correlation between the WGQ(6) and the SALSA (rho = 0.49), and we found a weak correlation between the WGQ(6) and the P-Scale (rho = 0.33). Both correlations are significant at the 0.01 level.

INTERNAL CONSISTENCY
A good Cronbach's alpha was found for the WGQ(11) (α = 0.81). An acceptable Cronbach's alpha was found for the WGQ(6) (α = 0.68).

FLOOR AND CEILING EFFECTS
No floor or ceiling effects were identified for the WGQ(11). Only 7 respondents (6.9%) scored the lowest possible score of 0, and nobody scored the highest possible score of 33.

Floor effects were found for the WGQ(6), as 17.8% (n=18) of the respondents scored the lowest possible score of 0. As nobody scored the highest possible score of 18, no ceiling effects were identified.

Variables		Spearman's rho
Construct validity	WGQ(11) – SALSA	0.64

	WGQ(11) – P-Scale Short	0.42	
	WGQ(6) – SALSA	0.49	
	WGQ(6) – P-Scale Short	0.33	
		Cronbach's alpha	
Internal consistency	WGQ(11)	0.81	
	WGQ(6)	0.68	
Floor & Ceiling effects	Floor effects	Yes, for the WGQ(6) (WGQ(6): 17,8%, WGQ(11): 6.9%)	
	Ceiling effects	No (WGQ(6): 0%, WGQ(11): 0%)	

Table 2. Quantitative variables WGQ(6) and WGQ(11)

The medians and inter-quartile range (IQR) per item are shown in Table 3. Participants had per item a possible score of 0-3. The median of most items (n=7) is 0, while the median of the other items (n=4) is 1. The IQR was mostly 1 (n=9). Items 2, 3 and 10 were best understood, as nobody needed an explanation or example. Question 7 was most difficult to understand, as 11 participants needed an example and/or explanation.

Item	Median	IQR	Explanation and/or example needed
1. Do you have difficulty seeing, even if wearing glasses?	0	1	1
2. Do you have difficulty hearing, even if using a hearing aid?	0	0	-
3. Do you have difficulty walking or climbing steps?	1	1	-
4. Do you have difficulty remembering or concentrating?	0	1	2
5. Do you have difficulty (with self-care such as) washing all over or dressing?	0	1	2
6. Using your usual (customary) language, do you have difficulty communicating, for example, understanding or being understood?	0	0	3
7. Do you have difficulty with activities or certain movements?	0	1	11
8. Do you experience numbness in your hands and/or feet?	1	1	1
9. Do you have weakness in your hands and/or feet?	1	1	1
10. Do you have wounds on your hands and/or feet?	1	1	-
11. Compared to others. do you have difficulty taking part in events, work, meetings, etc?	0	1	2

Table 3. Statistics per item WGQ scales: WGQ(6) contains questions 1-6, WGQ(11) contains all 11 questions (score: 0-3) (N=101). IQR: inter-quartile range.

To promote the interpretability of the scores in our study, we calculated the means, 95% confidence interval (CI 95%), medians and the interquartile range (IQR) of scores of different sub-groups. The data for the WGQ(6) and WGQ(11) is shown in Table 4 and Table 5, respectively.

To give an example how Table 4 and Table 5 can be interpreted in a qualitative way, Figure 4 shows the relation between participants' disability grade and their

mean total scores measured with the tools. The figures show that participants with a disability grade 1 or 2 have on the WGQ(11) a significant higher score than participants with a disability grade 0. This significance is not found for the WGQ(6). There is no significant difference in total scores when participants with a disability grade 1 and 2 are compared. Comparable findings are shown in the same figure (Figure 4) for the P-Scale Short and the SALSA Scale.

Variable		WGQ(6)			
		Mean	CI 95%	Median	IQR
	Entire sample (n = 101)	2,5	2.0 - 2.9	2	3
Gender	Male (n = 58)	2,4	1.8 - 3	2	2
	Female (n = 43)	2,6	1.9 - 3.3	2	3
Education	No education or did not finish primary school (n = 43)	3,2	2.5 - 3.9	3	4
	Education, at least finished primary school (n = 58)	1,9	1.4 - 2.4	1	2
Job	Yes, having a job* (n = 64)	2,1	1.6 - 2.6	2	2
	No, doesn't have a job** (n = 37)	3,1	2.3 - 4.0	2	3,5
Disability grade	Grade 0 (n = 43)	2	1.4 - 2.6	2	3
	Grade 1 (n = 22)	3,2	1.8 - 4.5	2	4,5
	Grade 2 (n = 36)	2,6	1.9 - 3.2	2	2,8

Table 4. Mean, CI 95%, median, IQR of WGQ(6) (total score: 0-18)

Variable		WGQ(11)			
		Mean	CI 95%	Median	IQR
	Entire sample (n = 101)	6,2	5.4 - 7.1	6	6
Gender	Male (n = 58)	6,3	5.1 - 7.5	6	6
	Female (n = 43)	6,2	4.8 - 7.5	6	6
Education	No education or did not finish primary school (n = 43)	7,1	5.7 - 8.6	6	6
	Education, at least finished primary school (n = 58)	5,6	4.5 - 6.7	5	6,3
Job	Yes, having a job* (n = 64)	5,3	4.4 - 6.3	4,5	5
	No, doesn't have a job** (n = 37)	7,8	6.1 - 9.4	7	9
Disability grade	Grade 0 (n = 43)	4,5	3.4 - 5.5	3	4
	Grade 1 (n = 22)	7,8	5.4 - 10.3	7	9,3
	Grade 2 (n = 36)	7,4	6.0 - 8.7	7	6,5

Table 5. Mean, CI 95%, median, IQR of WGQ(11) (total score: 0-33)
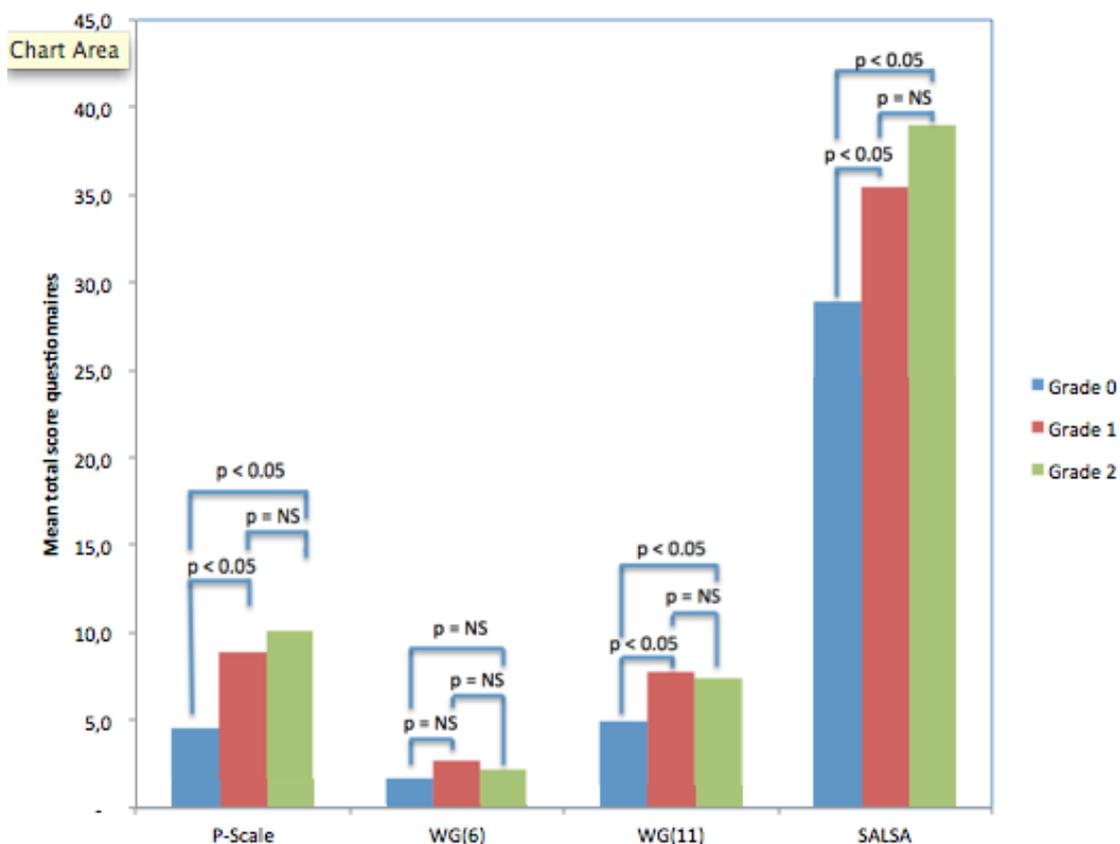
Figure 4. Mean total scores per scale (P-Scale Short; WGQ(6), WGQ(11); SALSA) per disability grade (grade 0; grade 1; grade 2). Participants with disability grade 1 or 2 have a significantly higher mean total score than participants with disability grade 0, when measured with the P-Scale Short, WGQ(11) or SALSA. p < 0.05 = significant. NS means not significant.

5.3.02 P-SCALE SHORT

Corresponding table of this section is Table 6.

CONSTRUCT VALIDITY

We found a moderate correlation between the P-Scale Short and the WGQ(11) (rho = 0.42) and between the P-Scale Short and the SALSA (rho = 0.48). We found a weak correlation between the P-Scale Short and the WGQ(6) (rho = 0.33). All correlations are significant at the 0.01 level (2-tailed).

INTERNAL CONSISTENCY

A good Cronbach's alpha (α = 0.72) was found for the P-Scale Short.

FLOOR AND CEILING EFFECTS

There are no floor effects, as 12.9% (n=13) respondents have the lowest possible score of 0. No ceiling effects are identified, as nobody scored the highest possible score of 70.

Variables		Spearman's rho
Construct validity	P-Scale Short – WGQ(11)	0.42
	P-Scale Short – WGQ(6)	0.33
	P-Scale Short – SALSA	0.48
		Cronbach's alpha
Internal consistency	P-Scale Short	0.72
Floor & Ceiling effects	Floor effects	No (12,9%)
	Ceiling effects	No (0%)

Table 6. Quantitative variables P-Scale Short

ITEMS

Every item is seen as relevant by more than 98% of the respondents. Two items were reported to be irrelevant for 2 out of 101 respondents, 7 items were irrelevant for 1 out of 101 respondents, and the other 4 items were seen as relevant by all respondents (see Table 7). An example of a reason why an item was reported as irrelevant, is a man who was living alone and did not have to discuss anything with somebody at home (item 12).

The medians and inter-quartile range (IQR) per item are shown in Table 7. Possible score per item is 0-5. The median of most items (n=12) is 0. Only one item has a median of 1. The IQR of the items is 0 (n=7), 1 (n=1) or 2 (n=5). Question 11 and question 13 were best understood, as nobody needed an explanation or example. Question 12 was most difficult to understand, as 15 participants needed an example and/or explanation.

Item	Median	IQR	Irrelevant	Prompt and/or example needed
1. Do you have equal opportunity as your peers to find work?	1	2	1	2
2. Do you work as hard as your peers do? (same hours, type of work etc.)	0	2	1	4
3. Do you contribute to the household economically in a similar way to your peers?	0	2	1	3
4. Do you make visits outside your village / neighbourhood as much as your peers do? (except for treatment) e.g. bazaars, markets	0	2	1	5
5. Do you take part in major festivals and rituals as your peers do? (e.g. weddings, funerals, religious festivals)	0	0	-	6
6. Do you take as much part in casual recreational/social activities as do your peers (e.g. sports, chat, meetings)	0	2	2	3
7. Do you have the same respect in the community as your peers?	0	0	-	1
8. Do you visit other people in the community as often as other people do?	0	0	-	7

Item	Median	IQR	Irrelevant	Prompt and/or example needed
9. Do you move around inside and outside the house and around the village / neighbourhood just as other people do?	0	0	1	2
10. In your village / neighbourhood, do you visit public places as often as other people do? (e.g. schools, shops, offices, market and tea/coffee shops)	0	0	2	2
11. In your home, do you do household work?	0	0	1	-
12. In family discussions, does your opinion count?	0	0	1	15
13. Are you comfortable meeting new people?	0	1	-	-

Table 7. Statistics per item P-Scale Short (score = 0-5) (N = 101)

INTERPRETABILITY

To promote the interpretability of the scores in our study, we calculated the means, 95% confidence interval (CI 95%), medians and the interquartile range (IQR) of scores of different sub-groups. The data for the P-Scale Short is shown in Table 8.

Variable		P-Scale Short			
		Mean	CI 95%	Median	IQR
	Entire sample (n = 101)	6.98	5.6 - 8.3	5	10
Gender	Male (n = 58)	5.7	4.1 -7.2	4	9
	Female (n = 43)	8.7	6.4 - 11.0	7	10
Education	No education or did not finish primary school (n = 43)	6.4	4.8 – 8.0	5	9
	Education, at least finished primary school (n = 58)	7.4	5.4 – 9.5	5	11
Job	Yes, has a job* (n = 64)	5.6	4.2 – 7.0	4	8
	No, does not have a job** (n = 37)	9.4	6.7 – 12.1	8	10
Disability grade	Grade 0 (n = 43)	3.7	2.6 - 4.9	3	5
	Grade 1 (n = 22)	8.5	5.3 - 11.7	6.5	12.3
	Grade 2 (n = 36)	9.9	7.3 - 12.6	10	10.8

Table 8. Interpretability (total score 0-65).
*Category 'no job' include answer options 'no work, because of health reasons', 'no work, because of other reasons', 'housewife', and 'voluntary'.
**Category job include answer options 'self-employee', 'government/private employee', and 'working on daily wages'.

5.3.03 SALSA

CONSTRUCT VALIDITY

We found a strong correlation between the SALSA and the WGQ(11) (Spearman's rho = 0.64) and a moderate correlation between the SALSA and the WGQ(6) (Spearman's rho = 0.49) and between the SALSA and P-Scale Short (Spearman's rho = 0.48). All correlations are significant at the 0.01 level (2-tailed).

A good Cronbach's alpha (α=0.89) was found for the SALSA.

FLOOR AND CEILING EFFECTS

There are no floor or ceiling effects identified, as nobody has the lowest or highest possible total score of respectively 0 and 80.

Variables		Spearman's rho
Construct validity	SALSA – WGQ(11)	0.64
	SALSA – WGQ(6)	0.49
	SALSA – P-Scale Short	0.48
		Cronbach's alpha
Internal consistency	SALSA	0.89
Floor & Ceiling effects	Floor effects	No (n=0)
	Ceiling effects	No (n=0)

Figure 5. Quantitative variables SALSA scale

ITEMS

The medians and the inter-quartile range (IQR) are per item shown in Table 9. Possible score per items is 0-4. The median per item is 1 (n=16) or 2 (n=4). The IQR per item is 0 (n=2), 1 (n=11), 2 (n=4) or 3 (n=3). Item 12 and question 18 were best understood, as nobody needed an explanation or example. Item 4 was most difficult to understand, as 9 participants needed an example and/or explanation.

Item	Median	IQR	Explanation and/or example needed
1. Can you see (enough to carry out your daily activities)**?**	1	1	2
2. Can you sit or squat on the ground?	1	1	1
3. Do you walk barefoot? e.g. most of the time	2	3	1
4. Do you walk on uneven ground?	2	2	9
5. Do you walk longer distances? i.e. longer than 30 minutes	2	2	5
6. Do you walk longer distances? i.e. longer than 30 minutes	1	0	7
7. Do you cut your finger or toenails? e.g. using scissors or clippers…	1	1	1
8. Do you hold a cup or basin with hot contents? e.g. drinks. food …	1	2	1
9. Do you work with tools? i.e. tools which you hold in your hands to help you work …	1	1	8
10. Do you carry heavy objects or bags? e.g. shopping. food. water, wood …	1	1	1
11. Do you lift object above your head? e.g. to place on a shelf, on your head, to hang clothes to dry …	1	1	6
12. Do you cook? i.e. prepare food both hot and cold	1	0	-
13. Do you pour hot liquids?	1	1	7
14. Do you open/close screw capped bottles? e.g. oil, water ..	1	1	8

Item	Median	IQR	Explanation and/or example needed
15. Do you open jars with screw--on lids? e.g. jam…	1	1	1
16. Do you handle or manipulate small objects? e.g. coins, nails, small screws, grains and seeds …	1	3	2
17. Do you use buttons? e.g. buttons on clothing, bags…	1	2	4
18. Do you thread needles? i.e. pass thread through the eye of a needle	2	3	-
19. Do you pick up pieces of paper, handle paper or put it in order?	1	1	5
20. Do you pick up things from the floor?	1	1	7

Table 9. Statistics per item SALSA scale (score: 0-4) (n=101).

INTERPRETABILITY

To promote the interpretability of the scores in our study, we calculated the means, 95% confidence interval (CI 95%), medians and the interquartile range (IQR) of scores of different sub-groups. The data for the SALSA scale is shown in Table 10.

Variable		SALSA Scale			
		Mean	CI 95%	Median	IQR
	Entire sample (n = 101)	33,2	31.0 - 35.5	31	17
Gender	Male (n = 58)	33,2	30.2 - 36.1	30	15,8
	Female (n = 43)	33,3	29.9 - 36.8	34	19
Education	No education or did not finish primary school (n= 43)	35,2	32.1 - 38.4	35	13
	Education, at least finished primary school (n = 58)	31,8	28.7 - 34.8	28	19,5
Job	Yes, having a job* (n = 64)	31,2	28.5 - 33.8	29	13
	No, doesn't have a job** (n = 37)	36,8	33.1 - 40.6	36,7	17,5
Disability grade	Grade 0 (n = 43)	27,3	24.5 - 30.2	24	10
	Grade 1 (n = 22)	35	29.3 - 40.8	35,5	25,5
	Grade 2 (n = 36)	39,2	36.3 - 42.1	38,5	12,5

Table 10. Mean, CI 95%, median and IQR of SALSA scale (total score: 0-80)

*Category 'no job' include answer options 'no work, because of health reasons', 'no work, because of other reasons', 'housewife', and 'voluntary'.

**Category job include answer options 'self-employee', 'government/private employee', and 'working on daily wages'.

6   DISCUSSION

This study aimed to contribute to the development of an internationally usable cross-NTD toolkit of instruments to measure NTD-related morbidity and disability by conducting a quantitative cultural validation study among persons affected by leprosy in West-Java, Indonesia. In this section, the results identified in this study are discussed per sub-question to become able to answer the main question.

6.1 CHARACTERISTICS STUDY POPULATION

This study included participants living in a former leprosy community: Sitanala Community, Bantam, West-Java, Indonesia. It is a heterogeneous study population, as people from different ages (29-99), gender (man/women), disability grades (0; 1; 2), educational levels (no education – finished high school), profession (no job – salaried employees) were included. The heterogeneity facilitates the generalizability of the study to other people affected by leprosy living in the same culture.

6.2 SUB-QUESTION 1: REGARDING WASHINGTON SET OF QUESTIONS

> ➢ To what extent are **the two versions of the WGQ** equivalent to the original version in terms of semantic, item, operational and measurement equivalence when used among persons affected by leprosy in West-Java, Indonesia?

6.2.01 SEMANTIC, ITEM AND OPERATIONAL EQUIVALENCE

During the pilot, the only item that needed to be revised was question 4, "Do you have troubles remembering or communicating with others due to your physical or mental condition?" No data is found why both the communication and cognition is asked in one question in the Indonesian version, while it is not this kind of double-barrelled question in other versions. We split this question into two questions: "Do you have difficulty remembering or concentrating?" and "Do you have difficulty understanding or being understood when communicating, using daily language?" Despite still a double-barrelled structure, this study shows no problems with these two new questions. Also an extensive qualitative validation study among several countries justifies keeping both aspects of the concepts (communication: understanding and being understood; cognition: remembering and concentrating) in one question per concept, like we did in this study (UN ESCAP, 2010). Therefore, it is recommended to use in the Bahasa versions the two new questions made in this study.

The average duration of 3.4 minutes to conduct the WGQ(11) is seen as acceptable. The average duration of the WGQ(6) was not measured, but will be

even shorter. A note has to be made that the measured time is only the time it took to ask and answer the tool's question. The actual time to conduct the questionnaire will be longer, as an introduction and explanation about how the tools works takes also time.

6.2.02 MEASUREMENT EQUIVALENCE

Until now, no other studies can be found with data about measurement properties of one of the WGQ versions. Therefore, comparing quantitative data found in this study with other data is not possible.

If we compare this study results of the WGQ(6) and WGQ(11), the WGQ(11) scores better on all tested psychometric properties. First, the data shows better construct validity for the WGQ(11) than the WGQ(6). A reason for this is that the WGQ(6) is less comprehensive in facing all aspects of the construct activity. In other words, the short WGQ(6) version lacks content validity, which affects the construct validity. Furthermore, this study shows good internal consistency for the WGQ(11) and an acceptable internal consistency for the WGQ(6). Last, both versions did not show ceiling effects. The WGQ(6) had floor effects, while the WGQ(11) had not. It shows the WGQ(11) is more sensitive in finding people facing activity limitations than the WGQ(6), which is acknowledged by the Washington Group (Madans & Loeb, 2011). Therefore, if the aim is to find all people with activity limitations, the WGQ(11) is more favourable.

Except question 7, all items were considered easy to understand. Question 7 "Do you have difficulty with activities or certain movements?" was found to be more difficult to answer, as it is a double-barrelled question and the terms "activities" and "certain movements" could be seen as broad and ambiguous. As in this study population it was common to be Muslim, the example of praying in Islam was given as "activity or certain movements", as it is an activity that includes several movements. Researchers who want to use this questionnaire should be aware of possible difficulties with this question and the context-specific example.

With calculating and discussing the interpretability, we want to show how the quantitative data can be used and interpreted in a qualitative way. This study shows a relation between participants' disability grade and their mean total score on the WGQ(11). Respondents with a disability grade 1 or 2 have significantly more activity limitations and participation restrictions than respondents with disability grade 0, when measured with the WGQ(11). This is not surprising when we look at the ICF model: body function & structure, activity and participation are all interrelated (WHO, 2001). The lack of a correlation between activity limitation measured with the WGQ(6) and disability grade might be explained by a lack of content validity, as we concluded earlier already as well. The general nature of the

37

activity questions in the WGQ(6) are not sufficiently sensitive to the type of neuro-disabilities assessed in the WHO disability grading system used in this study.

6.3 SUB-QUESTION 2: REGARDING P-SCALE SHORT

> To what extent is **the P-Scale Short** equivalent to the original version in terms of semantic, item, operational and measurement equivalence when used among persons affected by leprosy in West-Java, Indonesia?

6.3.01 SEMANTIC, ITEM AND OPERATIONAL EQUIVALENCE

The Bahasa Indonesia version of the P-Scale was already used in two large studies among persons affected by leprosy (n = 1299, n = 295), which were partly conducted on Java (Stevelink et al., 2012). Therefore, remaining problems with the translation were unlikely. Indeed, this study did not find problems regarding items or translation during the interviews: all items were considered to be relevant and acceptable to ask and everybody was able to understand the Bahasa language used in the interviews. The duration of on average 6.8 minutes to conduct the P-Scale Short was seen as acceptable.

6.3.02 MEASUREMENT EQUIVALENCE

As hypothesised, the P-Scale Short had moderate correlations with the WGQ(11) and SALSA Scale. Therefore, the construct validity is supported. Regarding internal consistency, Cronbach's alpha of 0.72 was found for the P-Scale Short. Originally, the authors of the P-Scale aimed for a combination of items with an alpha coefficient of ≥ 0.80 (Van Brakel et al., 2006). The authors of the shortened P-Scale found a Cronbach's alpha of 0.91 (Stevelink et al., 2012). Although the α in this study does not reach these levels (0.72), it still meets the criteria of adequate internal consistency (Terwee et al., 2007). This study shows no floor or ceiling effect for the P-Scale Short.

During the pilot, all items were considered as relevant and easy to understand. However, during analysis after conducting all interviews, a problem with semantic equivalence turned out: question 12 "In family discussions, does your opinion count?" needed often explanation. The difficulty was the word "count" as the original meaning was not transferred properly. The word is translated from English to Bahasa language in a word that implies calculation instead of importance. Therefore, in future research, this question should be reformulated. For example, "In family discussions, do people listen to your opinion?"
As it is for the WGQ(11), participants with a WHO disability grade 1 or 2 have more participation restriction than persons with disability grade 0, when

38

measured with the P-Scale Short. A reason for this is that persons affected by leprosy often face stigma and stigma is related with participation restriction (Peters, Lusli, & Miranda-Galarza, 2013; Sermrittirong & Van Brakel, 2014). Nevertheless, another study found no relation between participation and WHO disability grade when using the P-Scale (De Souza et al., 2016). A reason given by De Souza et al. (2016) is that the WHO grading system is based only on physical impairment, while participation is much more complex and involves also other aspects, such as environmental and psychosocial.

6.4 SUB-QUESTION 3: REGARDING SALSA SCALE

> To what extent is **the SALSA scale** equivalent to the original version in terms of semantic, item, operational and measurement equivalence when used among persons affected by leprosy in West-Java, Indonesia?

6.4.01 SEMANTIC, ITEM AND OPERATIONAL EQUIVALENCE

The answer option "If no, why not – I avoid because of risk" was deleted to avoid confusion among respondents, as researchers who are experienced with using the SALSA scale recommend this. A consequence of deleting the answer option "If no, why not? – I avoid because of risk" is that question 3 "Do you walk barefoot?" seem to be less relevant without this answer option. Another study showed that the answer option "If no, why not? – I avoid because of risk" is especially relevant for this question about walking barefoot (Wijk, Brandsma, Dahlström, & Björk, 2013). Therefore, it is recommended to remove this question if the answer option about risk avoidance is removed as well.

During the pilot study and training of interviewers, question 2 "*Do* you sit or squat on the ground" was changed into "*Can* you sit or squat on the ground?" During the interviews, no problems were mentioned about the *can* instead of *do* in the question. There were other reasons for confusion among this question. The question about both *sitting* and *squatting on the ground* led to difficulty, as people saw it as two different questions. People mentioned that sitting was easier than squatting on the ground. It would be preferable to split this question into two separate questions. For example: "Can you sit?" and "Can you squat on the ground". In an earlier Bahasa version of the SALSA scale, the question was already separated into two questions. For unknown reasons, researchers changed it back into one question again. We recommend to use two separate questions.

The duration of on average 5.9 minutes to conduct the SALSA scale was seen as acceptable.

6.4.02 MEASUREMENT EQUIVALENCE

The construct validity of the SALSA scale is supported, as this study's hypothesis is accepted. Internal consistency was very good, with a Cronbach's alpha of 0.89. This is almost the same as the internal consistency of the original SALSA scale (Cronbach's alpha = 0.88) (Velema, 2010). No floor or ceiling effects are found.

To improve interpretability, the relation between the mean total SALSA scores and participants' WHO disability grade was measured. This study found that participants with disability grade 1 or 2 faced more activity limitations than participants with grade 0. Other researchers found the same relation between WHO disability grades and activity limitation measured with the SALSA (De Souza et al., 2016). A reason is that SALSA activity questions are often directly related to specific structures and functions of the body, and these body structures and functions are taken into account with the WHO disability grade (De Souza et al., 2016).

6.5 LIMITATIONS

This study had several limitations that are worth mentioning.

There are little validation studies to compare this study's results with. It implies that this research gives valuable new information. However, it also made it difficult to compare this study's results with other studies.

A limitation of the study is that the interviews were conducted by 10 interviewers. If fewer interviewers would have done the interviews, possible inter-interviewer variability would have been lower. To minimise inter-interviewer variability, all except one interviewer received training in preparation for the interviews. The interviews done by the person who did not receive training were excluded from the analysis.

In this study, conceptual validity, semantic, and item equivalence were only examined in a pilot of 3 participants. A large qualitative study might have led to more in-depth insights of the respondents about the tools.

Due to time constraints, this study could not examine test-retest reliability of the instruments, while this is an essential part of measurement equivalence and cultural validity.

6.6 FURTHER RESEARCH

This study made an important step towards the development of an internationally usable cross-NTD toolkit of instruments to measure NTD-related morbidity and disability. More research in this field is needed, to validate the tools used in this study (WGQ(6); WGQ(11); P-Scale Short; SALSA) among persons affected with other NTDs and in other regions. Also other tools need to become validated on a large scale, to develop the aimed toolkit. During the same time as this study, colleague researches also validated these and other tools in several countries in order to develop an NTD-related morbidity and disability toolkit. It is recommended to compare data from these mentioned studies and this study to

get an overview of the limitations and strengths of the toolkit until so far. We hope that a well-validated set of instruments will make it possible to map NTD-related morbidity and disability in a way that promotes resource allocation and service provision to people who need help. Hopefully, one day, the term "neglected" of neglected tropical diseases will be a term from the past.

7   CONCLUSION

> ➢ Main question: How cultural valid are the two version of the WGQ, the P-Scale Short and the SALSA scale among persons affected by leprosy in West-Java, Indonesia?

Cultural validation of the tools[10] is in this study done regarding five equivalences: conceptual, semantic, item, operational and measurement equivalence. Measurement equivalence referred to the psychometric properties: construct validity, internal consistency, floor and ceiling effects, and interpretability.

The WGQ(6), which is the Washington set of questions consisting the 6 core questions, had adequate cultural equivalence for the conceptual, semantic, item and operational equivalency. Except for difficulties with item 7, the same can be concluded for the WGQ(11), which is an extended and more country- and disease-specific version. The measurement properties of the WGQ(11) were considered as being more valid than these of the WGQ(6). The WGQ(11) is more comprehensive in facing all aspects of the construct "activity", is more internal consistent, and is more sensitive in finding people facing activity limitations than the WGQ(6). Although the WGQ(11) is in these aspects favourable, the WGQ(6) has acceptable measurement properties as well. Concluding, the WGQ(6) was considered as a short, easy to use tool that gives an appropriate impression of the situation of activity limitation among persons affected by leprosy in West-Java. The WGQ(11) is more suitable if used in the study's context, but is less likely to be generalizable to other populations/regions, as it is a more country- and disease-specific tool.

In the P-Scale Short only question 12 "In family discussions, does your opinion count?" led often to confusion, due to the word "count". In future research, this question should be reformulated to, for example, "In family discussions, do people listen to your opinion?" Except for this, the conceptual, semantic, item, operational and measurement equivalence was comparable with the original versions and therefore supported by this study. Concluding, if item 12 is reformulated, the P-Scale Short is regarding a broad range of both qualitative and quantitative aspects suitable to use among persons affected by leprosy in West-Java.

This study showed problems with two items of the SALSA: item 2 and 3. We recommend to separate question 2 into two questions, and to delete question 3 if the answer option about risk avoidance is removed as well. Except for this, the conceptual, semantic, item, operational and measurement equivalence were

---

[10] All tools had some modifications during the pilot. It is recommended to use the new versions that include these modifications, if the tools are used in same context.

considered as good. Concluding, the SALSA scale is a suitable tool if question 2 and 3 are revised and when used among persons affected by leprosy in West-Java.

The overall conclusion is that this study shows that the WGQ(6); WGQ(11); P-Scale Short, SALSA are overall cultural valid to use among persons affected by leprosy in West-Java. This study results are not one-to-one generalizable to other NTDs or regions: carefully testing or validating in the new context is needed. Persons who want to use these and other tools need to be highly aware of the importance of cultural validation of tools. What works in one population, does not need to work in another. More validation on a broader scale will be needed to develop an internationally usable cross-NTD toolkit of instruments to measure NTD-related morbidity and disability. We are delighted that this study is another step towards the development of the toolkit.

BIBLIOGRAPHY

ASB Office for Indonesia. (2015). Use of the Washington Group questions on disability in Disaster Risk Reduction: An introduction. Retrieved April 14, 2016, from https://www.academia.edu/24971372/Use_of_the_Washington_Group_Questions_on_disability_in_Disaster_Risk_Reduction_An_introduction

Bowden, A., & Fox-Rushby, J. (2003). A systematic and critical review of the process of translation and adaptation of generic health-related quality of life measures in Africa, Asia, Eastern Europe. *Social Science & Medicine*.

De Souza, V., Da Silva, W., De Jesus, A., De Oliveira, D., Raptis, H., De Freitas, P., & Schneiberg, S. (2016). Is the WHO disability grading system for leprosy related to the level of functional activity and social participation? *Leprosy Review*, *87*(2), 191–200.

Ganihartono, I. (1996). Psychiatric morbidity among patients attending the Bangetayu community health centre in Indonesia. *Buletin Penelitian Kesehatan*.

Herdman, M., Fox-Rushby, J., & Badia, X. (1998). A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Quality of Life Research*.

Hotez, P., Molyneux, D., Fenwick, A., & Ottesen, E. (2006). Incorporating a rapid-impact package for neglected tropical diseases with programs for HIV/AIDS, tuberculosis, and malaria. *PLoS Med*.

Hotez, P., Ottesen, E., Fenwick, A., & Molyneux, D. (2006). The neglected tropical diseases: the ancient afflictions of stigma and poverty and the prospects for their control and elimination. *Hot Topics in Infection and Immunity in Children III*.

Irmansyah, I., Dharmono, S., Maramis, A., & Minas, H. (2010). Determinants of psychological morbidity in survivors of the earthquake and tsunami in Aceh and Nias. *Journal of Mental ....*

Kelders, R., Van Brakel, W., Beise, K., & Irwanto. (2012). Testing and validating a simplified scale to measure social participation of people with disabilities in Indonesia. *Disability and Rehabilitation*.

Kilham, S. (2015). Capacity building for disability inclusive disaster risk reduction in Indonesia Technical Report 2. Retrieved April 14, 2016, from http://sydney.edu.au/health-sciences/cdrp/projects/pipddmi.shtml

Kirshner, B., & Guyatt, G. (1985). A methodological framework for assessing health indices. *Journal of Chronic Diseases*.

Kolaczinski, J., Kabatereine, N., & Onapa, A. (2007). Neglected tropical diseases in Uganda: the prospect and challenge of integrated control. *Trends in Parasitology*.

Kostanjsek, N. (2011). Use of The International Classification of Functioning, Disability and Health (ICF) as a conceptual framework and common language for disability statistics and. *BMC Public Health*.

Kowal, P., Kahn, K., Ng, N., Naidoo, N., Abdullah, S., Bawah, A., ... Tollman, S. (2010). Ageing and adult health status in eight lower-income countries: the indepth WHO-SAGE collaboration. *Global Health Action*.

Kurniawan, A., & Meyliandrie, L. (2014). Self-Regulation of Indonesian Red Cross Volunteer Correlated to Big 5 Personality. Retrieved April 14, 2016, from http://iafor.info/archives/offprints/acss2014-offprints/ACSS2014_0749.pdf

Lohr, K., Aaronson, N., Alonso, J., Burnam, M., Patrick, D., Perrin, E., & Roberts, J. (1996). Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clinical Therapeutics*.

Madans, J., & Loeb, M. (2011). Measuring disability and monitoring the UN Convention on the Rights of Persons with Disabilities: the work of the Washington Group on Disability Statistics. *BMC Public*.

Melchior, H., & Velema, J. (2011). A comparison of the Screening Activity Limitation and Safety Awareness (SALSA) scale to objective hand function assessments. *Disability and Rehabilitation*.

Ng, N., Hakimi, M., Byass, P., Wilopo, S., & Wall, S. (2010). Health and quality of life among older rural people in Purworejo District, Indonesia. *Global Health Action*.

Peters, R., Lusli, M., & Miranda-Galarza, B. (2013). The meaning of leprosy and everyday experiences: an exploration in Cirebon, Indonesia. *Journal of Tropical*.

Peters, R., Van Brakel, W., Zweekhorst, M., Damayanti, R., Bunders, J., & Irwanto. (2014). The cultural validation of two scales to assess social stigma in leprosy. *PLoS Negl Trop*.

SALSA Collaborative Study Group. (2007). The development of a short questionnaire for screening of activity limitation and safety awareness (SALSA) in clients affected by leprosy or diabetes. *Disability and Rehabilitation*.

Sermrittirong, S., & Van Brakel, W. H. (2014). Stigma in leprosy : concepts , causes and determinants. *Leprosy Review*.

Stevelink, S., Hoekstra, T., Nardi, S., Van der Zee, C., Banstolaf, N., Premkumarg, R., ... Van Brakel, W. (2012). Development and structural validation of a shortened version of the Participation Scale. *Disability and Rehabilitation*.

Stevelink, S., & Van Brakel, W. (2013). The cross-cultural equivalence of participation instruments: a systematic review. *Disability and Rehabilitation*.

Tan, M., Kusriastuti, R., Savioli, L., & Hotez, P. (2014). Indonesia: an emerging market economy beset by neglected tropical diseases (NTDs). *PLoS Negl Trop Dis*.

Terwee, C., Bot, S., De Boer, M., Van der Windt, D., Knol, D., Dekker, J., ... De Vet, H. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*.

UN. (2006). Convention on the Rights of Persons with Disabilities. Retrieved April 13, 2016, from http://www.un.org/disabilities/documents/convention/convoptprot-e.pdf

UN ESCAP. (2010). *Results of the Testing of the ESCAP/WG Extended Question Set on Disability*.

Üstün, T. (2010). Measuring health and disability: Manual for WHO disability assessment schedule WHODAS 2.0.

Van 't Noordende, A., Kuiper, H., Ramos, A. N., Mierasa, L. F., Barbosab, J. C., Pessoab, S. M. F., ... Van Brakel, W. H. (2016). Towards a toolkit for cross-neglected tropical disease morbidity and disability assessment. *International Health*.

Van Brakel, W., Anderson, A., Mutatkarc, R., Bakirtziefd, Z., Nichollse, P., Rajuf, M., & Das-Pattanayakg, R. (2006). The Participation Scale: measuring a key concept in public health. *Disability and Rehabilitation*.

Van Brakel, W., Sihombing, B., Djarir, H., Beise, K., Kusumawardhani, L., Yulihane, R., ... Wilder-Smith, A. (2012). Disability in people affected by leprosy: the role of impairment, activity, social participation, stigma and discrimination. *Global Health*.

Van Veen, N., Hemo, D., Bowers, R., Pahanb, D., Negrinib, J., Velemac, J., & Richardusa, J. (2011). Evaluation of activity limitation and social participation, and the effects of reconstructive surgery in people with disability due to leprosy: a prospective cohort study. *Disability and Rehabilitation*.

Velema, J. (2010). SALSA Scale, Users Manual. Retrieved February 21, 2016, from https://www.leprosy-information.org/files/SALSA manual v1.pdf

WHO. (2001). International classification of functioning, disability and health: ICF. Retrieved April 4, 2016, from https://books.google.nl/books?hl=nl&lr=&id=lMZPmEJrJ3sC&oi=fnd&pg=PA1&dq=ICF+model+2001+who&ots=PFlN_B9aDe&sig=7-mSPAGjS_YThZoh0zl3eNVW8SI

WHO. (2010). Working to overcome the global impact of neglected tropical diseases: first WHO report on neglected tropical diseases. Retrieved April 8, 2016, from http://apps.who.int/iris/handle/10665/44440

WHO. (2011). The Global Network for Neglected Tropical Diseases Control (GNNTDC). Retrieved April 8, 2016, from http://www.who.int/neglected_diseases/GNNTDC/en/

WHO. (2013). Morbidity management and disability prevention in lymphatic filariasis. Retrieved April 13, 2016, from http://apps.searo.who.int/pds_docs/B4990.pdf

WHO. (2014). Metrics: Disability-Adjusted Life Year (DALY). Retrieved April 13, 2016, from http://www.who.int/healthinfo/global_burden_disease/metrics_daly/en/

WHO. (2015). Neglected tropical diseases. Retrieved April 8, 2016, from http://www.who.int/neglected_diseases/diseases/en/

Wijk, U., Brandsma, J., Dahlström, O., & Björk, M. (2013). The concurrent validity of the Amharic version of Screening of Activity Limitation and Safety Awareness (SALSA) in persons affected by leprosy. *Leprosy Review*.

8 APPENDIX

8.1 POSSIBLE RELEVANT TOOLS

ICF Domain	Tool	Used/Validated in Indonesia	Validation needed on leprosy	Bahasa version available	Remarks
Body functions and struc-tures	Clinical Profile*	No	Yes	No	New tool: made and validated only by Van 't Noordende et al. (2016). Translation is needed first.
	SRQ*	Yes. For example: (Ganihartono, 1996; Irmansyah, Dharmono, Maramis, & Minas, 2010; Kurniawan & Meyliandrie, 2014)	Yes	Yes	Screens for problems in mental health.
Activity	WHODAS 2.0*	Yes. (Kowal et al., 2010; Ng, Hakimi, Byass, Wilopo, & Wall, 2010; Üstün, 2010)	Yes	Yes	A study in Sulawesi showed that the WHODAS 2.0 was not easy to use among people with low literacy
	SALSA	Yes. (Van Brakel et al., 2012)	Yes	Yes	Studies showed the SALSA could also be used for measuring activity limitations in persons affected by leprosy in Ethiopia and Bangladesh (Melchior & Velema, 2011; Van Veen et al., 2011; Wijk et al., 2013).
	WGQ	Yes. (ASB Office for Indonesia, 2015; Kilham, 2015; Madans & Loeb, 2011)	Yes	Yes	Potential to become widely used worldwide, as it is promised as the main disability screening tool (Madans & Loeb, 2011)
Partici-pation	P-Scale Short*	Yes. (Kelders, Van Brakel, Beise, & Irwanto, 2012; Stevelink et al., 2012)	Yes	Yes	Shortened version of the P-Scale, while all items of the original P-scale are still covered (Stevelink et al., 2012)
Environ-mental factors	CHIEF*	No	Yes	No	Translation is needed first.
	EMIC-CSS*	Yes. (Peters et al., 2014)	No	Yes	Need for another sample, as this tool focuses on community members, while the others focus on persons affected by NTDs themselves (Peters et al., 2014).
Personal factors	WHOQOL-BREF*	Yes. (Manuscript submitted)	Yes	Yes	No results available yet.
	WHOQOL-DIS*	No	Yes	No	During this study, colleague R. M. Brunnekreeft translated this tool to Bahasa

* Tool that is included in the NMD toolkit of Van 't Noordende et al. (2016)

8.2 INTERVIEW GUIDES

8.2.01 DEMOGRAPHIC FEATURES

Question	Answering options
1. Record sex as observed	Female
	Male
2. How old are you now?	__ Years
3. How many years in all did you spend studying in school, college or university?	__ Years
4. What is your current marital status? (Select the single best option)	Never married
	Currently married
	Separated
	Divorced
	Widowed
	Cohabiting
5. Which describes your main work status best? (Select the single best option)	Paid work
	Self-employed, such as own your own business or farming
	Non-paid work, such as volunteer or charity
	Student
	Keeping house/ homemaker
	Unemployed (health reasons)
	Unemployed (other reasons)
	Other (specify)

8.2.02 PILOT INTERVIEWS

TO BRING ALONG
Interpreter, interview guide, questionnaire, pen, recorder, informed consent, notebook, gift.

PART I: INTERVIEW INTRODUCTION
- Thank for participation of participant
- Introduce who we are
- Introduce the underlying research and the coming interview
- Explain:
    - Participant can take as much time as he/she thinks is needed;
    - Participant may always ask questions if something is not clear;
    - It is okay if a participant don't want to answer a question;
    - We use the data carefully;
    - The answers are anonymously used; no one will be able to find out which participant gave which answers.
- Ask for permission to record the interview
- Ask for acceptation of the informed consent; written or verbal.

- Record begin of the application of the instrument
- Apply the instrument
- Register during the interviews the following aspects:
  - Questions for which an **e**xample needs to be given before the question is understood (mark with '**E**')
  - Questions that need to be **r**eformulated before the question is understood (mark with '**R**')
  - Questions that are understood the first time the question is asked (mark with '√')
  - Questions that are not answered ('**X**')
- Stop recording time when last question is answered

PART III: INTERVIEW QUESTIONS

Evaluate the instrument. Note that these questions are about the application of the instrument, not about the time and conversation before and after application.
- What was your overall impression of this interview?
  - Why?
- Were the questions relevant to you?
  - Why (not)?
- Do you think that all questions were acceptable to ask?
- Did you find the duration of the set of questions acceptable?
  - Why (not)?
- Why did you need an example/reformulation for question X?
  - Do you have suggestions how this question could be better formulated?
- Is there anything else you would like to tell us?

PART IV: FINISHING INTERVIEW
- Thank for participation of participant

8.2.03 QUANTITATIVE

TO BRING ALONG
Interpreter, interview guide, questionnaire, pen, recorder, informed consent, notebook, gift.

PART I: INTERVIEW INTRODUCTION
- Thank for participation of participant
- Introduce who we are
- Introduce the underlying research and the coming interview
- Explain:
  - Participant can take as much time as he/she thinks is needed;
  - Participant may always ask questions if something is not clear;

- o It is okay if a participant don't want to answer a question;
- o We use the data carefully;
- o The answers are anonymously used; no one will be able to find out which participant gave which answers.
- Ask for permission to record the interview
- Ask for acceptation of the informed consent; written or verbal.

PART II: APPLYING INSTRUMENT X
- Record begin of the application of the instrument
- Apply the instrument
- Register during the interviews the following aspects:
  - o Questions for which an **e**xample needs to be given before the question is understood (mark with '**E**')
  - o Questions that need to be **r**eformulated before the question is understood (mark with '**R**')
  - o Questions that are understood the first time the question is asked (mark with '√')
  - o Questions that are not answered ('**X**')
- Stop recording time when last question is answered

PART III: SHORT EVALUATION
Evaluate the instrument. Note that these questions are about the application of the instrument, not about the time and conversation before and after application.
- What was your overall impression of this interview?
  - o Why?
- Is there anything else you would like to tell us?

PART IV: FINISHING INTERVIEW
- Thank for participation of participant

8.3 INFORMED CONSENT

Principle investigator: Liesbeth Akkerman
Name of organization: Netherlands Leprosy Relief / Universitas of Indonesia (UI)
Title of the study: Towards developing a cross-neglected tropical disease (NTD) toolkit for assessment and monitoring of NTD-related morbidity and disability

INTRODUCTION

Thank you very much for your time. We are Liesbeth and Arafat and we work for NLR and Universitas of Indonesia (UI). We want to find more information about the situation of persons affected by leprosy on Java. To do this, we use questionnaires. Your opinion is important to see if the questionnaires are understandable and usable. If there is something you do not understand, please ask me to stop and I will take time to explain. If you have questions later, you can ask them at any time to me or the interviewer.

VOLUNTARY PARTICIPATION

Your participation in this research is voluntary. You always have a choice to participate or not. You don't need to give a reason why you want to stop. If you do or do not participate, does not influence your treatment. There is no direct benefit for you to participate in this research. Hopefully, our study will help to map the situation of people affected by leprosy in several countries.

PROCEDURES

First, Arafat will ask you the questions of the first questionnaire. You can take as much time for the questions as you need. After the first questionnaire, Arafat will ask you opinion about the questions. Again, you can take as much time as you prefer. Then, Arafat asks the questions of and about the second questionnaire. And then of the third questionnaire. In total, it will take about one hour. The questions of the questionnaire are personal questions. If you don't want to answer a question, you can skip the question and we will continue to the next question. We will record the conversation. This record is highly confidential. Only we have access the information you gave. We don't talk with other about the answers you gave. Do you have questions about what I explained? I will answer your questions.

CONSENT OF PARTICIPANTS

We explained you the topic of this study. You had the opportunity to ask questions and any questions you asked are answered to your satisfaction. You had the opportunity to refuse to participate in this study. You consent voluntarily to be a participant in this study. Do you agree with this?

Name of participant: Date: Signature:

8.4 WASHINGTON SET OF QUESTIONS

Question	Original WGQ (used in this study as WGQ(6))	Questionnaire used by Van Brakel et al., 2012	Extended WGQ version (used in this study as WGQ(11))
Do you have difficulty with seeing, even when wearing glasses?	X	X	X
Do you have difficulty hearing, even if using a hearing aid?	X		X
Do you have difficulty walking or climbing steps?	X	X	X
Do you have difficulty remembering or concentrating?	X		X
Do you have difficulty washing all over or dressing?	X	X	X
Using your usual (customary) language, do you have difficulty communicating, for example understanding or being understood?	X		X
Are there activities that you cannot perform?		X	X
Do you have loss of feeling in the hands and/or feet?		X	X
Do you have any weakness in hands and/or feet?		X	X
Do you have any wounds on hands and/or feet?		X	X
Do you have any wounds on hands and/or feet?		X	X
Do you have any problems in relationships or in taking part in festivities, work, meetings, etc?		X	X

Table 11. Questions included in the original and extended version of the WGQ: WGQ(6) and WGQ(11)

NOTE: Each WGQ question has four response categories: no – no difficulty; yes – some difficulty; yes – a lot of difficulty; and cannot do at all. The responses scales are scored from 0-4, where "no – no difficulty" is scored as 0 and "cannot do at all" as 4.

8.5 P-SCALE SHORT

No Participation Scale-Short	Not specified, not answered	Yes	Sometimes	No	Irrelevant, I don't want to, don't have to	NO problem	Small	Medium	Large SCORE
1. Do you have equal opportunity as your peers to find work?		0			0				
[if sometimes or no] How big a problem is it to you?						1	2	3	5
2. Do you work as hard as your peers do? (same hours, type of work etc)		0			0				
[if sometimes or no] How big a problem is it to you?						1	2	3	5
3. Do you contribute to the household economically in a similar way to your peers?		0			0				
[if sometimes or no] How big a problem is it to you?						1	2	3	5
4. Do you make visits outside your village / neighbourhood as much as your peers do? (except for treatment) e.g. bazaars, markets		0			0				
[if sometimes or no] How big a problem is it to you?						1	2	3	5
5. Do you take part in major festivals and rituals as your peers do? (e.g. weddings, funerals, religious festivals)		0			0				
[if sometimes or no] How big a problem is it to you?						1	2	3	5
6. Do you take part in social activities as much as your peers do? (e.g. sports, chat, meetings, religious, or community activities)		0			0				
[if sometimes or no] How big a problem is it to you?						1	2	3	5
7. Do you have the same respect in the community as your peers?		0			0				
[if sometimes or no] How big a problem is it to you?						1	2	3	5
8. Do you visit other people in the community as often as other people do?		0			0				
[if sometimes or no] How big a problem is it for you?						1	2	3	5
9. Do you move around inside and outside the house and around the village / neighbourhood just as other people do?		0			0				
10. In your village / neighbourhood, do you visit public places as often as other people do? (e.g. schools, shops, offices, market and tea/ coffee shops)		0			0				
[if sometimes or no] How big a problem is it to you?						1	2	3	5
11. In your home, do you do household work?		0			0				
[if sometimes or no] How big a problem is it to you?						1	2	3	5
12. In family discussions, does your opinion count?	0				0				
[if sometimes or no] How big a problem is it to you?						1	2	3	5
13. Are you comfortable meeting new people?	0			0					
[if sometimes or no] How big a problem is it to you?						1	2	3	5
									TOTAL

8.6 SALSA SCALE

	SALSA scale **S**creening of **A**ctivity **L**imitation & **S**afety **A**wareness **Tick one box on each line in** **response to each question.**	If Yes, how easy is it for you?			If not, why not?	
		Easy	**A Little difficult**	**Very difficult**	**I don't need to do this**	**I physically cannot**
1	**Can you see** (enough to carry out your daily activities)**?**	1	2	3		4
2	**Do you sit or squat on the ground?**	1	2	3	0	4
3	**Do you walk barefoot?** e.g. most of the time	1	2	3	0	4
4	**Do you walk on uneven ground?**	1	2	3	0	4
5	**Do you walk longer distances?** i.e. longer than 30 minutes	1	2	3	0	4
6	**Do you wash your whole body?** (using soap, sponge, jug;	1	2	3	0	4
7	**Do you cut your finger or toenails?** e.g. using scissors or clippers…	1	2	3	0	4
8	**Do you hold a cup or basin with <u>hot </u>contents?** e.g. drinks,	1	2	3	0	4
9	**Do you work with tools?** i.e. tools which you hold in your hands to help you work …	1	2	3	0	4
10	**Do you carry heavy objects or bags?** e.g. shopping, food, water, wood …	1	2	3	0	4
11	**Do you lift objects above your head?** e.g. to place on a shelf, on your head, to hang clothes to dry …	1	2	3	0	4
12	**Do you cook?** i.e. prepare food both hot and cold	1	2	3	0	4
13	**Do you pour hot liquids?**	1	2	3	0	4
14	**Do you open/close screw capped bottles?** e.g. oil, water ..	1	2	3	0	4
15	**Do you open jars with screw-on lids?**	1	2	3	0	4
16	**Do you handle or manipulate small objects?** e.g. coins, nails, small screws, grains and seeds …	1	2	3	0	4
17	**Do you use buttons?** e.g. buttons on clothing, bags…	1	2	3	0	4
18	**Do you thread needles?** i.e. pass thread through the eye of a needle	1	2	3	0	4
19	**Do you pick up pieces of paper, handle paper or put it in order?**	1	2	3	0	4
20	**Do you pick up things from the floor?**	1	2	3	0	4