WORLD BANK GROUP

Report No: AUS0003146

Funded by the European Union via the Structural Reform Support Programme and implemented by the World Bank, in cooperation with the European Commission

РЕПУБЛИКА БЪЛГАРИЯ
МИНИСТЕРСТВО НА ТРУДА И СОЦИАЛНАТА ПОЛИТИКА

Bulgaria

Options for Including Functioning into Disability Status Assessment

Carolina Fellinghauer, Aleksandra Posarac, Jerome Bickenbach and Marijana Jasarevic

Report No: AUS0003146

Bulgaria: Options for Including Functioning into Disability Status Assessment

Carolina Fellinghauer, Aleksandra Posarac, Jerome Bickenbach, and Marijana Jasarevic

2022

1818 H Street NW, Washington DC 20433
Telephone: 202-473-1000; Internet: www.worldbank.org

Table of Contents

List of Tables

List of Figures

Acknowledgements

This Report was prepared by a World Bank Team led by Marijana Jasarevic, Social Protection Specialist of the World Bank. The team included Carolina Fellinghauer, Lead Statistical Analyst, University of Zurich, Department of Psychology, Chair for Psychological Methods, Evaluation and Statistics, Aleksandra Posarac, Lead Economist and Lead Technical Advisor to the Project, Professor Jerome Bickenbach, Lead Technical Expert, Swiss Paraplegic Society and University of Luzern and Viktorija Vasiljeva-Gringiene, WHODAS Pilot Operational Coordinator, Violeta Nikolova and Eleonora Petrova, WHODAS Pilot Field Coordinators and IT expert Vesela Minkovska. The Team is grateful for close collaboration and assistance from: Lyubomira Arshinkova, State Expert, Department for Integration of People with Disabilities within the Directorate Disability Policy, Equal Opportunities and Social Assistance at the Ministry of Labor and Social Policy, Marin Baychev, Head, Department for Integration of People with Disabilities Directorate Disability Policy, Equal Opportunities and Social Assistance at the Ministry of Labor and Social Policy, Marc Vothknecht, Head of the Sector for Labor Market and Social Polices, The EC Directorate-General for Structural Reform Support (DG REFORM), Claudia Piferi, Policy Officer, DG REFORM, Oana Dumitrescu, Policy Officer, DG REFORM. Sylvia Stoyanova, Operations Officer, World Bank Country Office Bulgaria, Adela Ivanova Delcheva Nachkova, Operations Assistant. The team would also like to thank Cem Mete, Social Protection and Labor Practice Manger and Fabrizio Zarcone, and Lasse Melgaard, World Bank Resident Representatives in Bulgaria for the overall support and guidance provided to the team.

An Overview

1. Background

This Note was prepared as part of the project "Strengthening Disability System in Bulgaria" implemented by the World Bank with funding from, and in collaboration with, the European Commission's Directorate General for Structural Reform Support - DG REFORM. The specific objective of the project is to support the Ministry of Labor and Social Policy (MLSP) of Bulgaria to strengthen and further develop its disability system, including by supporting the MLSP in strengthening the individual comprehensive assessment of functioning and needs of persons with disabilities and related administrative processes.

To advise on the further development of the assessment of needs of persons formally certified as having a disability of at least 50 percent by Territorial Medical Expert Medical Commissions (TMEC) or National Medical Expert Commission (NMEC),[1] it was necessary to evaluate to what extent the current system considers functioning in the assessment of disability. The comprehensive individual needs assessment in Bulgaria, as regulated by the Persons with Disabilities Act (adopted in December 2018, in force since January 1, 2019, and subsequent amendments) and operationalized by the Social Assistance Agency (SAA) of MLSP is based on functioning as understood by the International Classification of Functioning, Disability and Health of the World Health Organization (WHO ICF, 2001).[2] All persons with disabilities certified in Bulgaria to have a disability are required to undergo the individual needs assessment to access benefits targeted at persons with disabilities. The certificate of disability serves as a basis on which the needs assessment is conducted. It is therefore very important that these two processes are harmonized in their respective approaches to disability.

To that end, to evaluate disability status assessment from the perspective of functioning, the World Health Organization's tool for measurement of disability – Disability Assessment Schedule – WHODAS was pilot-tested to collect information from 3,118 persons who underwent disability assessment at the end of 2021 - beginning of 2022. The details on the pilot are provided in the Bulgaria WHODAS Pilot Protocol, which is attached to this Note (Attachment 1).

This Note presents the results of the statistical analysis of the pilot data set and recommendations on options for including functioning into disability status assessment. A comprehensive description and review of the current disability status assessment and the individual needs assessment systems can be found Chapters 3 and 4, respectively of the *Bulgaria: Disability System and Policy, A Comprehensive Review*.[3]

[1] For the purpose of the PDA, persons with disabilities are understood as "persons with permanent physical, mental, intellectual and sensory impairment who may impede their full and effective participation in public life and to whom the medical expertise has established a degree of disability of 50 and over 50 percent".[1] In Article 101, The Health Act stipulates that medical expertise is conducted to establish "temporary work incapacity, type and degree of disability of children up to 16 years of age and of persons who have acquired the right to a social insurance pension based on age and length of work history covered by social insurance contributions according to Article 68 of the Social Insurance Code, and to establish a degree of permanently reduced of working age adults, as well as to confirm the presence of professional disease.[1] In this Note, reflecting the PDA, we use the term persons with disabilities.

[2] World Health Organization. 2001. *International Classification of Functioning, Disability and Health (ICF)*. https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health.

[3] Posarac, A. et al. 2022, *Bulgaria: Disability System and Policy, A Comprehensive Revie*w. © World Bank.

2. About the World Health Organization's Disability Assessment Schedule (WHODAS)

In the ICF, information about categories of Activities and Participation can be collected either from the ***perspective of capacity*** (reflecting exclusively the expected ability of a person to perform activities considering their health conditions and impairments) or the ***perspective of performance*** (reflecting the actual performance of activities in the real-world environmental circumstances in which the person lives). Information about capacity typically represents the results of a clinical inference or judgment based on medical information, while performance is a true description of what occurs in a person's life. The two perspectives are therefore very different, although capacity constitutes a determinant of performance.

A disability assessment is a summary measure of the level of a person's performance of an adequately representative set of behaviors and actions, simple to complex, in their actual environment, considering the person's state of health.

The WHO developed, tested and has consistently recommended the WHODAS as an instrument that can validly and reliably capture the performance of activities by an individual in his or her daily lives and actual environment. The 'actual environment' is represented in the ICF in terms of environmental factors that act either as environmental facilitators (e.g., assistive devices, supports, home modifications) or as environmental barriers (inaccessible houses, streets and public buildings, stigma, and discrimination). The WHODAS questionnaire, in short, is WHO's recommended, generic, performance-based disability assessment tool.

3. Implementing WHODAS in Bulgaria

WHODAS 36-question version was implemented in Bulgaria on a sample of 3,118 individuals who applied for disability (re)assessment in late 2021 and early 2022. The pilot sample included only persons who were assessed as having a disability of at least 50.0 percent. The survey was conducted in collaboration with the Social Assistance Agency of MLSP and more than 60 social workers participated as interviewers, while day to day pilot monitoring was conducted by the two pilot coordinators. Because of the social distancing restrictions in Bulgaria at the time of the pilot, only 66.7 percent of the interviews took place face-to-face, while 33.3 percent were phone interviews. The pilot sample included more female than male applicants (53.5 percent vs. 46.5 percent respectively). The average age was 56.2 years. A little over half of the applicants were currently married (50.7 percent); 12.4 percent were widowed; 11.9 percent were divorced, and 4.7 percent were cohabiting. Most applicants were living independently in the community (99.3 percent). The applicants had an average of 11.7 years of education. Most applicants reported either being unemployed for health reasons (35.4 percent) or being retired (26.4 percent). Only 30.8 percent reported having a paid employment. All applicants reported one primary ICD-10 linked health condition with additional comorbidities. Neoplasms (23.9 percent) and diseases of the circulatory system (23.0 percent) were the most reported main diagnoses. Mental and behavioral disorders were reported by 11.3 percent of applicants. ICD chapter XIII (diseases of the musculoskeletal system and connective tissue) and ICD chapter IV (endocrine, nutritional, and metabolic diseases) were the primary diagnoses in 8.1 percent and 7.7 percent, respectively.

4. Data analysis

Below, we summarize the results of the pilot data set data analysis. First, we analyze metric and psychometric properties of the WHODAS pilot data and then we look at the current disability assessment methodology outcomes of the pilot participants as compared to the WHODAS assessment.

a. Psychometric properties of WHODAS 2.0 in Bulgaria

A statistical analysis of psychometric properties of WHODAS pilot that included seven essential statistical tests (described in the main text below) show that the data collected with WHODAS, under the Rasch analysis, display robust psychometric properties of validity and reliability. It is important to keep in mind that the WHO developed WHODAS explicitly to statistically capture the construct of functioning from the perspective of performance – namely the experience of performing activities by a person with an underlying health problem in their actual everyday life environment. There is an abundance of evidence from the scientific literature – supported by the results of this pilot – that WHODAS is a psychometrically sound instrument that reliably and validly collects information about levels of disability.

Therefore, we can confidently conclude that information collected with the WHODAS is robust, viable and relevant and that it validly represents the construct of disability as understood in ICF and the Convention on the Rights of Rights of Persons with Disabilities (CRPD). It can, thus, surely be included into the disability status assessment in Bulgaria to: (i) significantly strengthen the method of assessment currently in use (a medical assessment mostly based on impairments); (ii) bring it closer to the ICF and CRPD understanding of disability; and (iii) harmonize the approach to assessment with the ICF functioning based approach used in the individual needs assessment.

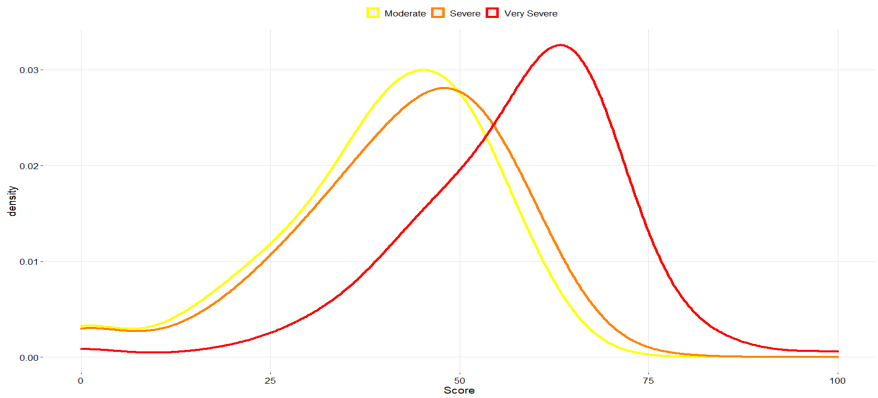
b. Comparing WHODAS and certified disability degree data

One of the objectives of our analysis of the Bulgaria WHODAS collected data is to show that the inclusion of functioning into the current medically based disability assessment method will significantly improve its capacity to assess the experience of disability more accurately and to allow for better assessment of needs of persons with disabilities subsequently.

The WHODAS data set included not only WHODAS collected data, but also the certified disability degree as determined by TMECs/NMEC and associated ICD codes for each participant in the WHODAS survey. This allowed us to compare the two sets of data.

The current disability assessment method in Bulgaria is medically based and uses an instrument of a Baremic type) that matches diseases and associated impairments with percentages of disability (as compared to a healthy person) . Procedurally, the assessment is based on medical documentation, justifying the degree of impairments in diseased or injured body part or structure, detailed relevant clinical history, in-depth clinical examination, and, in some cases, targeted laboratory and examinations performed by TMEC/NMEC. The current Bulgarian disability assessment system identifies disability degrees as percentages – with values < 50 percent designating mild/no disability, 50-70 percent moderate disability, 70-90 percent severe disability, and > 90 percent very severe disability. In what follows we will call these here 'disability severity ranking groups'.

Figure 1: WHODAS score density line by disability severity



Looking at the WHODAS functioning score by current Bulgarian disability severity ranking groups, it is observed that the medical assessment does not differentiate well between moderate and severe disability, suggesting low reliability and precision. The match is stronger in the case of very severe disability. Figure 1 shows that while the WHODAS scores for very severe functioning restrictions stand out (red line), the difference between severe and moderate disability severity (the yellow and orange lines) is less obvious with a closer location to each other. The density lines in Figure 1 also suggest the presence of false positives (high disability percentage and low WHODAS score) and false negatives (lower disability percentage and high WHODAS score). A more accurate assessment would show the very severe WHODAS density line sloping more to the right-hand side: the line would be closer to 0 up until the score of 45, then sharply rising around the score of 50. The opposite should be the case for the moderate disability, which should be located mostly to the left-hand side of the Figure 1. This suggests that the medical information may misrepresent the true extent of individual disability as experienced in daily life.

As noted above, that the assessment method used in Bulgaria is based on medical conditions and associated impairments and pre-determined percentages of disability – the so called Baremic method.[4] It is significant that concerns about this method are widely held in both the scientific and policy communities. A good and clear example of these concerns in the specific case of disability assessment is a Council of Europe's report on disability assessment in Europe, published in 2002.[5] In this report, the Baremic method is characterized as "an arbitrary ordinal scale which attaches progressive percentage values to define disabilities. The disabilities of the claimant are compared to those for which there are scale values, and a percentage is thereby obtained".

Thus, the results presented above come as no surprise as WHODAS was designed explicitly to assess whole- person disability, while the medical approach to assessing disability used in Bulgaria, does not directly assess disability, but infers disability based on the underlying health condition or impairment. Sometimes there is a very close correlation between severity of health condition and severity of disability; but sometimes there is no connection. We clearly see this in the case of mental health problems where the impact of the person's environment may greatly increase the impact of the experience of, e.g., depression. This is the basic validity problem with medically based disability

[4] The method of transposing such scales to percentages was introduced by the French mathematician François Barrême in late XVIII Century.

[5] Council of Europe. 2002. *Assessing Disability in Europe, Similarities and Differences*, Report drawn up by the Working Group on the assessment of person-related criteria for allowances and personal assistance for people with disabilities (Partial Agreement) (P-RR-ECA). Integration of people with disabilities. https://rm.coe.int/16805a2a27/.

assessment. As pointed out above, although the presence of a health condition and associated impairments is a precondition for disability, inferring the level of disability from the presence of a health condition is scientifically problematic. The level of disability that an individual experiences, as the ICF argues, is determined by an interaction between a health condition and associated impairments and environment in which the person lives. WHODAS was designed to directly capture this disability experience while assessment of disability based solely on medical grounds cannot do so validly or reliably.

c. Real life examples

To illustrate the discussion above, we present 6 randomly selected six real life cases from the WHODAS pilot data set where disability percentage and WHODAS scores differ dramatically (also summarized in Table 1).

Case A is a 62-year-old divorced man with a WHODAS-based functioning score of 63, which would indicate very severe functioning restrictions. He has been determined having a moderate disability severity of 66 percent. He reports 11 years of education and lives independently in the community but cannot work for health reasons. His main condition is an unspecified cirrhosis of liver, but he further presents a personality disorder and hypertensive heart disease. He reports also having had difficulties because of his health condition on every day of the last month. He is unable to perform usual activities and often must reduce usual activities or work.

Case B is a 59-year-old married woman with 15 years of education. She is currently working. She suffers an unspecified cirrhosis of liver accompanied by some anemia, a hip arthrosis, and gout. Her disability has been rated as very severe, i.e., 93 percent. Her WHODAS-score of 36 indicates that she has only moderate functioning problems in a day-to-day life. She also reports marginal difficulties in carrying out her activities and work, having to reduce or cut-back activities only about one day per month, when she is not feeling too well.

Case C is an 84-year-old married man with a WHODAS-based functioning score of 79. He has been determined having a 'severe' disability with a percentage of 72 due to a Type 2 Diabetes and a heart failure. He is retired but still lives independently in the community. His health condition is severely limiting him in his daily life, and he cannot perform his usual activities normally without having to reduce them.

Case D is a 45-year-old educated and working woman. She has never been married. She was diagnosed with diabetes mellitus with multiple complications, including hypertension and chronic pancreas problems. Her work reduction capacity has been rated as very severe, i.e., 94 percent. Based on the ratings of the WHODAS, she reports only moderate functioning problems in daily life, her score being 27. She also reports only marginal difficulties in carrying out her activities and work, she is never totally unable to carry out her work or activities because of her conditions and only must slow down somewhat from time to time.

Case E is a 44-year-old married woman with 17 years of education. She is unemployed but not for health reasons. She is in the severe disability severity group with a percentage of 85. She has been diagnosed with an organic personality disorder without further comorbidities. Her WHODAS-score of 22 indicates good functioning and her health condition is not limiting him in performing daily life activities.

Case F is a 19-year-old married man with 12 years of education living independently in the community. He is unemployed for health reasons. He is in the moderate disability severity group with a percentage of 50. He has been diagnosed with a mild mental retardation without further comorbidities. His

WHODAS-score of 60 is high and indicates severe functioning problems with his health condition limiting him every day of the month, having to reduce activities half of the time.

Table 1: Disability percentages and WHODAS scores and severity grouping - real life cases

Case	Disability % and group	WHODAS score and group
Case A	66% - moderate	63 – very severe
Case B	93% - very severe	36 - moderate
Case C	72% - severe	79 – very severe
Case D	94% - very severe	27 - moderate
Case E	85% - very severe	22 – no difficulty
Case F	50% - moderate	60 – very severe

The cases presented above corroborate the above discussion that that the assessment based on medical information may misrepresent the true extent of disability an individual experiences. This is very important, because an accurate assessment of disability is crucial for persons experiencing disability to access disability benefits. For example, cases A, C and F will have no access to personal assistants (their disability percentage is less than the threshold of 90 percent), although they experience severe difficulties in functioning. In contrast, cases B and D will be eligible for personal assistance, although their disability experience in terms of functioning is moderate. Including functioning into disability assessment in Bulgaria will, thus, not only improve the accuracy of the extent of disability assessment but will also improve the assessment of needs of persons with disabilities.

d. To conclude

The empirical evidence presented in the sections above shows that:

- WHODAS – is a freely available and widely used questionnaire built on the activity and participation domains of the WHO's International Classification of Functioning, and Health, that is as close to being the gold standard for the description of disability as we have. It is psychometrically strong, and the data can be analyzed to create a valid and reliable interval-scaled functioning score. This evidence from the Bulgarian pilot corroborates evidence from other international research studies: WHODAS successfully collects functioning information, and as it has been further confirmed by pilot data, it does so with strong psychometric properties of validity and reliability. It performs well in measuring whole-person disability, creates a summary score, and provides an objective and accurate assessment of functioning based on core functioning domains of the ICF. The scores provide interval-scales values ranging normally from 0 to 100 (Figure 8).
- The current system that determines disability severity ranking groups in Bulgaria exhibits well-known scientific concerns about its capacity to validly assess the true extent of disability an individual experiences and it does not differentiate degrees of disability accurately. The percentage continuum from 0 to 100 is poorly populated and polarizes on a few values.
- In light of these results from the pilot, we conclude that including the assessment of functioning based on WHODAS[6] into the assessment already in place in Bulgaria will significantly improve the

[6] We recommend WHODAS, because it is free and firmly empirically proven that it represents the construct of disability in terms of ICF and is psychometrically valid and reliable. Countries may choose to develop their own instruments, but such effort requires time and money, and the instrument will have to be psychometrically tested before being deployed.

accuracy of the assessment, resulting in a more refined assessment that adds information on the lived experience of the disability to the assessment based on the medical diagnosis.

5. Options for including functioning into disability assessment in Bulgaria

a. Methodological considerations

The WHODAS pilot in Bulgaria has shown that it performs well in capturing disability experience. The question is how best to include the functioning information captured by WHODAS into the disability status assessment system in Bulgaria.

It should be emphasized that we are not suggesting that medical information, or even assessment of kind and degree of disability based on medical information, should not play a role in disability assessment in Bulgaria. This is clearly not the case. The ICF itself makes it clear that without an underlying health condition and associated impairments, disability does not exist, so medical information will also be relevant to disability assessment. While we, and the scientific community, find problematic the direct inference of whole person disability from medical information alone, this information must still be collected and relied on for disability assessment. Information about health states provide a basis for identifying specific physical and mental dimensions of activities and areas of participation that are vulnerable to disability, which can then be directly confirmed by WHODAS data. Medical information provides essential guidance on the medium- and long-term trajectory of disability that the individual will experience, including whether the person faces a progressive decline in health capacity resulting in more and more disability, or the reverse, a progressive improvement. In short, medical information is an essential component of disability assessment.

As medical information is essential, in this section of the Report, we analyze and discuss possible options for combining medical and functioning information in the assessment of disability in Bulgaria.

As we have done in other countries, several methods were tested on the Bulgarian pilot dataset to address this challenge. These methods can be grouped here into two principal strategies (1) ***averaging*** the medical assessment percentage with the WHODAS score to arrive at a final disability assessment score, and (2) ***flagging*** persons whose WHODAS score, and disability severity group are different from the severity group based on the percentage determined based on medical information.[7]

(1) **Averaging** – averaging the attributed disability percentage and WHODAS score. This approach is based on the theory that, together, medical, and functioning scores contribute, to different degrees, to a realistic and valid assessment of disability. In the main text below, we describe the results of four strategies that were tested using different weighting combinations.

(2) **Flagging** – identifying persons whose WHODAS severity grouping differs from the medically determined severity grouping and flagging these individuals to request from them additional information or reassessment. When an individual has a WHODAS score over or below some cut-off, this suggests that the medical score does not adequately capture the experience of disability and a second level assessment should be conducted.

Averaging and Flagging are, arguably, the most intuitively obvious approaches to merging diverse assessments into a single overall assessment. Each is grounded in the ICF understanding of disability as the outcome of an interaction between the underlying health condition and impairments of a

[7] It is important to add that as WHODAS is used more data are collected, this data can be further analyzed using the techniques from this Note to continually update and recalibrate parameters and cut points. Moreover, these data have other potential policy applications, including in identifying disability trends and planning for the future.

person and the physical, human-built, interpersonal, attitudinal, social, economic, and political environment in which the person lives and acts. They differ, however, in how they weigh the impact of the medical and environmental determinants of disability.

b. Options for including the assessment of functioning into disability assessment in Bulgaria

Below, we present options to include functioning into disability assessment in Bulgaria. Each option follows the ICF theory in as much as it combines the medical component of assessment with a functioning component, assessed by WHODAS. Option A is the situation in which WHODAS scores are considered in a purely discretionary manner. Options B (averaging strategies), and C (flagging strategies) are quantitative. Each of these options has advantages and disadvantages. Our framework for evaluating them – based on the scientific literature – are key scientific principles that determine the credibility of any disability assessment process: validity (the extent to which the option relies on a true assessment of disability); reliability (the ability of the option to arrive at the same assessment of the same case by different assessors); transparency (the degree to which the assessment process and outcomes can be described and understood by all stakeholders); and standardization (the extent to which the process resists distortion or alteration over time and across locations).

Option A: Discretionary combination of medical and functioning components

This is the option in which an individual or committee reviews medical scores and the WHODAS scores and makes a judgment about the extent of disability as the individual or committee sees fit. This is a purely discretionary option, and it is surprisingly common in practice. This approach is subject to manipulation, or whim, lacks validity and reliability, and is utterly non-transparent. The option is given here as a contrast to the remaining options B and C, but also, in fairness, because some countries continue to rely on this option for disability assessment. We do not recommend this option.[8]

Options B and C: quantitative approach

Averaging and Flagging options are quantitatively driven, which makes them very different from Option A. In different ways and for different reasons, they satisfy not only the basic psychometric properties of validity and reliability but each, to different degrees, strives to achieve transparency and standardization.

Option B: Using an averaging algorithm

In the Bulgaria pilot WHODAS data set, there is a relatively high percentage of persons indicating no functioning problems at all (10.7 percent), among which some individuals were in the very severe disability severity ranking group. Averaging the disability percentages with the WHODAS score would adjust the number of persons in each of the disability severity ranking groups by accounting to some degree for the observed disability level assessed by the WHODAS. To get a full sense of the range of possible approaches under Option B, four weighting schemes were tested: (i) 75.0 percent disability percentage & 25.0 percent WHODAS score; (ii) 50.0 percent disability percentage & 50.0 percent WHODAS score; 25.0 percent disability percentage & 75.0 percent WHODAS score; and 0.0 percent disability percentage & 100.0 percent WHODAS score.

[8] Numerous interactions with officers involved in disability assessment in different countries suggest that medical professionals involved in the assessment disability are confident they "know best" and can consider functioning and the experience of disability as part of the medical description of the applicant's situation. One often hears medical assessors claim that they take functioning fully into account when examining medical records. One implicit result from the pilot is that this assumption is not grounded in evidence.

***Advantages of Option* B:** (i) An assessment of the level of functioning plays a significant role in the determination of eligibility for disability benefits so that the eligibility for benefits is not solely based on purely medical criteria. (ii) The averaging approach minimalizes the impact of the inherent psychometric problems with the disability percentage based on the Baremic medical assessment used. (iii) The assessment of the level of functioning is empirically and statistically verified. (iv) This option yields high levels of validity and reliability. (v) Merging the results of two assessments scaled by means of 'weighted averaging' is fully objective, transparent, and non-discretionary. (vi) The method is not sample-dependent.

***Disadvantages of Option B*: (i)** There are, potentially, an infinite number of combinations of weighting schemes (i.e., 'strategies'), each of which affects the set of eligible applicants differently and has different budgetary and political consequences. This is an unavoidable fact about the nature of disability as a continuum and the fact that there are not yet scientifically verified or objective cut-offs for severity on a Rasch scale 0-100 continuum. (ii) Any strategy selected will be objectionable to individuals who, under that strategy, will not be certified as disabled and thus not eligible for any benefits. This signals the need for clear and transparent information dissemination and a solid grievance redress system that may include using tools for clinical testing and determination of functioning.

Option C: Using the flagging algorithm

We tested three flagging strategies. Flagging persons with severe to very severe functioning problems whose disability severity was certified as moderate would result in many individuals (437 persons) whose disability severity in terms of functioning would have to be reassessed. Only flagging those with very severe functioning problems (WHODAS score of at least 60) who are in the moderate or severe disability ranking groups results in only 31 persons whose disability severity would need to be reassessed and augmented. On the other hand, a relatively large number of individuals (N = 217) in the severe and very severe disability ranking groups presented no disability in terms of the WHODAS scores. From those, almost half had neoplasms as main diagnosis. This reiterates the point raised above that the current disability assessment method does not discriminate well between different degrees of disability.

***Advantages of Option C*:** (i) Scientifically robust and based on actual data. (ii) Shows that the purely medical approach to disability assessment may not accurately assess disability in many cases – in which, as reported in the WHODAS score, a person is experiencing more/fewer functioning problems in their lives than what the health condition/impairment is thought to imply. (iii) High levels of validity and reliability.

***Disadvantages of Option C*:** (i) The WHODAS cut-offs for different degrees of functioning problems were recommendations based on past pilots and some evidence from the scientific literature. Sensitivity analyses are not available to this point. More precise cut-values specific to Bulgaria may be introduced at later time points when more information on functioning is collected (assuming that WHODAS will be introduced in Bulgaria). (ii) Technically robust methodological and procedural instructions will have to be developed to guide the reassessment process to ensure transparency.

Even with the caveat concerning the cut-off points for disability severity, the flagging method may be introduced through specifically designed (two step) administrative procedure.

Table 2 gives an overview of the testing strategies that were considered and gives the number of individuals that would be considered having a moderate, severe, or very severe disability after adjusting for the WHODAS-score. Further, the number of individuals that would have their disability severity ranking group changed towards a higher or a lower group are shown.

Table 2: Overview of WHODAS inclusion strategies

General Approach	Nbr.	Description of scores integration formula	Cut-off	Total Moderate Disability	Total Severe Disability	Total Very Severe Disability	Total Upshift Disability	Total Downshift Disability
Actual approach	#1	Reduced Working Capacity [%Disability*]	• No disability < 50% • Moderate 50-70% • Severe 70-90% • Very Severe > 90%	1096	912	1110	0	0
Averaging:	#2	Weighted mean of %Disability (75%) and WHODAS (25%)	Bivariate cut-offs for %Disability cut-offs and equivalent critical WHODAS Score level	1235	938	945	30	334
	#3	Weighted mean of %Disability (50%) and WHODAS (50%)		1342	998	778	74	652
	#4	Weighted mean of %Disability (25%) and WHODAS (75%)		1346	1080	692	224	892
	#5	Weighted mean of %Disability (0%) and WHODAS (100%)		1315	1133	670	552	1211
Flagging:	#6	WHODAS-Score = Severe disability & %Disability = Moderate		1053	955	1110	437	0
	#7	WHODAS-Score = Ver severe disability & %Disability = Moderate or severe		1090	893	1135	31	0
	#8	WHODAS-score = No to mild disability and %Disability = Severe or very severe		1258	805	1055	0	217

**% Disability stands for the medically attributed disability percentage or work capacity reduction percentage*

To make the options concrete, we illustrate them on the 6 real life cases presented above. This should show how the strategies would change the understanding of the level of disability and highlight the advantage to including functioning into the current disability assessment in Bulgaria. Table 3 presents the expected level of disability given each of the functioning inclusion strategies (yellow = moderate; orange = severe; red = very severe).

Table 3: Disability severity ranking and WHODAS scores and their integration strategies – Examples of individual cases

			Current method severity	Averaging				Flagging		
	WHODAS score	**Disability percent**	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**
A	63	66						**Additional information and second step assessment**		
B	36	93								
C	79	72								
D	27	94								
E	22	85								
F	60	50								

6. Implementation considerations

This Note has provided evidence that the current disability assessment system in Bulgaria would benefit significantly from the inclusion of functioning into the assessment method: (i) the assessment of disability would be more precise and accurate, reflecting the real life experience of disability of applicants; (ii) the assessment will be in line with modern understanding of disability; and (iii) the assessment will be harmonized with the individual needs assessment providing valuable input into it.

Our approach to disability assessment is to combine medical and functioning information and we have provided above several methodological options for doing it. An important further question is whether Bulgaria has administrative capacity to implement the change smoothly and without significant cost. The answer is a sound “yes”.

First, Bulgaria has an advanced information system that could easily accommodate the collection and use of the information on functioning.

Second, Bulgaria has a cadre of experienced social workers in the Social Assistance Agency that could be engaged in the WHODAS administration. While administrative process will have to be designed and details worked out, it could possibly flow in the following way: a person applying for/referred to the assessment of disability would have two meetings scheduled: one with the social worker to administer WHODAS and subsequently one with the TMEC. The WHODAS information would be sent to NMEC electronically where the form would be checked, and the raw score transformed into the Rasch based score. TMEC will proceed with the assessment as per the current criteria.

How the two scores will then be combined depends on the choice made by the Government. If the averaging method is chosen, say with 50.0 percent weight given to the TMEC determined degree of disability and 50.0 percent to the WHODAS Rasch score, the two scores will automatically be combined at NMEC, and the final score sent to the TMEC to issue the certificate. The certificate can also be issued by NMEC. If the flagging method is chosen, then in cases where the TMEC determined percentage of

disability and the WHODAS score fall in the same disability grouping (no, moderate, severe, and very severe), the NMEC will instruct the TMEC to issue the certificate with the proposed disability severity grouping. If they do not coincide, then a secondary assessment is undertaken either by a different TMEC or a NMEC. Whichever the ultimate choice might be, the result is that the information on functioning will be systematically included in disability assessment using a standardized approach, and the administrative process itself will become more rigorous, standardized, and objective.

Finally, it should also be noted that any new method adopted should apply to new applicants only. To smooth the transition, disability recertification may be staged over several years.

7. Recommendations

Based on the above and not to repeat what has already been said, we strongly recommend that the ***Bulgarian Government includes functioning into disability assessment using WHODAS to collect relevant information***. While the choice is political, either averaging or the flagging approach can comfortably be implemented based on the existing information systems and human resources (a cadre of social workers).

Including functioning into disability assessment will:

- make the assessment of disability more precise, accurate and reliable, reflecting the real-life experience of disability of applicants,
- bring the assessment closer to modern understanding of disability as formulated by ICF and mandated by CRPD, and
- align it with the individual needs assessment by providing valuable information input into it. A status assessment that includes functioning will provide a better profile of disability that the person experiences to identify needs that, once addressed, will improve the experience of disability by optimize the person's functioning.

We also recommend that a separate assessment tool is developed for children because the tools used for adults are not suitable for children. (WHO does not recommend that WHODAS is used for children and is currently working on a WHODAS instrument for children).

Bulgaria: Options for Introducing Functioning into Disability Status Assessment

1. Introduction

This Note was prepared as part of the project "Strengthening Disability System in Bulgaria" (EC reference SRSP2020/49 (20BG06) implemented by the World Bank with funding from, and in collaboration with, the European Commission's Directorate General for Structural Reform Support - DG REFORM. The specific objective of the project is to support the Ministry of Labor and Social Policy (MLSP) of Bulgaria to strengthen and further develop its disability system, namely by supporting the MLSP in strengthening the individual comprehensive assessment of functioning and needs of persons with disabilities and related administrative processes and supporting MLSP in the development of the institutional and governance structure of the newly proposed State Agency for People with Disabilities. The project expected outcomes are: (i) The State Disability Agency effectively coordinates and monitors disability policy in Bulgaria; and (ii) The Ministry of Labor and Social Policy carries out a well-preforming comprehensive assessment of functioning and needs of persons with disabilities. These outcomes are also in line with the goals of implementation of the government policy on the rights of persons with disabilities. The expected outcomes support the national priorities to guarantee effective social inclusion to people with disabilities.

To advise on the further development of the assessment of needs of persons formally certified as having a disability of at least 50 percent by Territorial Medical Expert Medical Commissions (TMEC) or National Medical Expert Commission (NMEC), it was necessary to evaluate to what extent the current disability assessment system considers the level of the disabled individual's functioning in the assessment. The comprehensive individual needs assessment in Bulgaria, as regulated by the Persons with Disabilities Act (2018) and operationalized by the Social Assistance Agency (SAA) of MLSP is based on functioning as understood by the International Classification of Functioning, Disability and Health of the World Health Organization (WHO ICF, 2001).[9] All persons with disabilities certified in Bulgaria as having a disability are required to undergo the individual needs assessment to access benefits targeted at persons with disabilities. The certificate of disability serves as a basis on which the needs assessment is conducted. It is therefore important that these two processes are harmonized in their approaches to disability.

To that end, to evaluate the disability status assessment from the perspective of functioning, the World Health Organization's tool for measurement of disability – Disability Assessment Schedule – WHODAS was pilot-tested to collect information from more than 3,100 persons who underwent disability assessment at the end of 2021 – beginning of 2022. Only persons who were assessed as having a degree of disability of at least 50.0 percent were included in the pilot. The details on the pilot are provided in the Bulgaria WHODAS Pilot Protocol, which is attached to this Note (Attachment 1).

This Note presents the results of the statistical analysis and recommendations on options for including functioning into disability status assessment. A comprehensive description and review of the current disability status assessment and the individual needs assessment systems can be found Chapters 3 and 4, respectively of the *Bulgaria: Disability System and Policy, A Comprehensive Review*.[10]

[9] World Health Organization. 2001. International Classification of Functioning, Disability and Health (ICF). https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health.

[10] A. Posarac et al, 2022, *Bulgaria: Disability System and Policy, A Comprehensive Revie*w. © World Bank

2. About the World Health Organization's Disability Assessment Schedule (WHODAS)

In the ICF, information about categories of Activities and Participation can be collected either from the perspective of *capacity* (reflecting exclusively the expected ability of a person to perform activities in light of their health conditions and impairments, without the influence of environmental facilitators or barriers) or the perspective of *performance* (reflecting the actual performance of activities in the real-world environmental circumstances in which the person lives). Information about capacity typically represents the results of a clinical inference or judgment based on medical information, examination, and observation, while performance is a true description of what occurs in a person's life. The two perspectives are therefore very different, although capacity constitutes a determinant of performance.

As an administrative act of formally establishing disability, the assessment of disability should be based on the overall lived experience of an individual living with one or more health problems – or in ICF terms, it is the level of a person's performance in light of their intrinsic health capacity and the impact of environmental facilitators or barriers that should be assessed. *Disability assessment is a 'whole person' or global assessment of the extent or level of a person's experience of disability*. This is important because disability assessment should be a summary measure of functioning levels across domains of actions, simple and complex, from walking, taking care of children to working at a job. A summary or global assessment of disability, of necessity, must be based both on the individual health state and on assessments of specific activities. Yet a summary assessment of disability is valid only if the specific assessments can be statistically summarized into a single assessment score.

The ICF understands a disability to be any level of problem or difficulty a person experiences in functioning in some domain, from the perspective of performance. The WHO developed, tested, and has consistently recommended the WHO Disability Assessment Schedule (WHODAS) as a questionnaire to capture the performance of activities by an adult individual in his or her daily life and actual environment. The 'actual environment' is represented in the ICF in terms of environmental factors that act either as environmental facilitators (e.g., assistive devices, supports, home modifications) or as environmental barriers (e.g., inaccessible houses, streets, and public buildings, stigma, and discrimination). The WHODAS questionnaire, in short, is WHO's recommended, generic, performance-based disability assessment tool. It is structured around six basic functioning domains of Activities and Participation (denoted in the ICF as D Codes):

- D1: Cognition – understanding & communicating
- D2: Mobility– moving & getting around
- D3: Self-care– hygiene, dressing, eating & staying alone
- D4: Getting along– interacting with other people
- D5: Life activities– domestic responsibilities, leisure, work & school
- D6: Participation– joining in community activities

Table 4: WHODAS items for the 36-item long form

	In the past 30 days, how much difficulty did you have in:
	Understanding and communicating
D1.1	Concentrating on doing something for ten minutes?
D1.2	Remembering to do important things?
D1.3	Analyzing and finding solutions to problems in day-to-day life?
D1.4	Learning a new task, for example, learning how to get to a new place?
D1.5	Generally understanding what people say?
D1.6	Starting and maintaining a conversation?
	Getting around
D2.1	Standing for long periods such as 30 minutes?
D2.2	Standing up from sitting down?
D2.3	Moving around inside your home?
D2.4	Getting out of your home?
D2.5	Walking a long distance such as a kilometer [or equivalent]?
	Self-care
D3.1	Washing your whole body?
D3.2	Getting dressed?
D3.3	Eating?
D3.4	Staying by yourself for a few days?
	Getting along with people
D4.1	Dealing with people you do not know?
D4.2	Maintaining a friendship?
D4.3	Getting along with people who are close to you?
D4.4	Making new friends?
D4.5	Sexual activities?
	Life activities
D5.1	Taking care of your household responsibilities?
D5.2	Doing most important household tasks well?
D5.3	Getting all the household work done that you needed to do?
D5.4	Getting your household work done as quickly as needed?
D5.5	Your day-to-day work/school?
D5.6	Doing your most important work/school tasks well?
D5.7	Getting all the work done that you need to do?
D5.8	Getting your work done as quickly as needed?
	Participation in society:
D6.1	How much of a problem did you have in joining in community activities in the same way as anyone else can?
D6.2	How much of a problem did you have because of barriers or hindrances in the world around you?
D6.3	How much of a problem did you have living with dignity because of the attitudes and actions of others?
D6.4	How much time did you spend on your health condition or its consequences?
D6.5	How much have you been emotionally affected by your health condition?
D6.6	How much has your health been a drain on the financial resources of you or your family?
D6.7	How much of a problem did your family have because of your health problems?
D6.8	How much of a problem did you have in doing things by yourself for relaxation or pleasure?

The professionally administered (i.e., "clinical") version of the WHODAS questionnaire collects information about functioning and problems in functioning – i.e., disability – by means of an interview conducted by a trained interviewer who asks standardized questions – and if necessary, follow-up probe questions. Considering the responses the respondent provides, the interviewer then makes a judgment about the level of disability experienced and uses WHODAS's 5-level response scale (1 = None, 2 = Mild, 3 = Moderate, 4 = Severe, 5 = Extreme or Cannot do) to rate and record answers to each question in the WHODAS for that individual. It should be reiterated that, as used in this pilot, WHODAS is not being used as a self-report questionnaire; it is rather being used as a questionnaire administered in face-to-face or telephone interviews by a trained professional. Respondents are informed that their answers about each domain of functioning should adopt the perspective of performance – that is, they should describe what they do in their daily lives, considering all environmental barriers and facilitators that they experience. The WHODAS 36-item, "clinically" administered version was chosen for the pilot to collect information about a representative range of functioning domains to create a full and complete picture of the overall level of disability experienced by the respondents in their everyday life. The 36 items are shown in Table 4 by functioning domain.

3. Descriptive statistics from the WHODAS pilot data

A pilot sample included a total of 3,118 individuals who applied for disability assessment in late 2021 and early 2022. The interviews were conducted by more than 60 social workers from the Social Assistance Agency of MLSP. Because of the social distancing restrictions in Bulgaria at the time of the pilot, only 66.7 percent of the interviews took place face-to-face, while 33.3 percent were phone interviews. Table 5 presents descriptive statistics of the assessed population. The pilot sample included more female applicants (53.5 percent vs. 46.5 percent respectively). The average age was 56.18 years (SD = 13.68). A little over half of the applicants were currently married (50.7 percent); 12.4 percent were widowed; 11.9 percent were divorced, and 4.7 percent were cohabiting. Most applicants were living independently in the community (99.3 percent). The applicants had an average of 11.7 (SD = 3.38) years of education. Most applicants reported either being unemployed for health reasons (35.4 percent) or being retired (26.4 percent). Only 30.8 percent reported having a paid employment.

All applicants reported one primary ICD-10[11] linked health condition with additional comorbidities (N = 3,118, 100.0 percent). Table 6 presents the frequency and percentages of observed ICD-10 diagnostic chapters for the applicants' primary health condition. Neoplasms (N = 758; 23.95 percent) and diseases of the circulatory system (N = 727; 22.97 percent) were the most reported main diagnoses. Mental and behavioral disorders were reported by N = 357 (11.28 percent) applicants. ICD chapter XIII (diseases of the musculoskeletal system and connective tissue) and ICD chapter IV (endocrine, nutritional, and metabolic diseases) were the primary diagnoses in N = 255; 8.06 percent and N = 245; 7.74 percent, respectively.

[11] WHO International Classification of Diseases, 10th revision.

Table 5: WHODAS Pilot Sample Descriptive Statistics

Pilot Size	3 118
Gender: Male (%)	1 449 (46.5)
Age - mean (SD)	56.18 (13.68)
Years of Education - mean (SD)	11.71 (3.38)
Marital Status (%)	
Never married	587 (18.8)
Currently married	1 580 (50.7)
Separated	44 (1.4)
Divorced	372 (11.9)
Widowed	388 (12.4)
Cohabiting	147 (4.7)
Living Condition (%)	
Independent in the community	3 095 (99.3)
Assisted living	15 (0.5)
Hospitalized	8 (0.3)
Work Status (%)	
Paid work	945 (30.8)
Self-employed	79 (2.6)
Non-paid work	0 (0.0)
Student	12 (0.4)
Keeping house	10 (0.3)
Retired	811 (26.4)
Unemployed (health reasons)	1 089 (35.4)
Unemployed (other reasons)	124 (4.0)
Other	3 (0.1)

Table 6: Prevalence of Health conditions in the study population

by ICD-10 Health Condition Category

ICD-Chapter	N	%
I Certain infectious and parasitic diseases	11	0.35 %
II Neoplasms	758	23.95 %
III Diseases of the blood	14	0.44 %
IV Endocrine, nutritional, and metabolic diseases	245	7.74 %
IX Diseases of the circulatory system	727	22.97 %
V Mental and behavioral disorders	357	11.28 %
VI Diseases of the nervous system	204	6.45 %
VII Diseases of the eye and adnexa	161	5.09 %
VIII Disease of the ear and mastoid process	37	1.17 %
X Diseases of the respiratory system	44	1.39 %
XI Diseases of the digestive system	53	1.67 %
XII Diseases of the skin and the subcutaneous tissue	22	0.7 %
XIII Diseases of the musculoskeletal system and connective tissue	255	8.06 %
XIV Diseases of the genitourinary system	60	1.9 %
XIX Injury, poisoning, and certain other consequences of external causes	97	3.06 %
XVII Congenital malformations, deformations, and chromosomal abnormalities	25	0.79 %
XVIII Symptoms, signs, and abnormal clinical and laboratory findings	1	0.03 %
XXI External causes of morbidity and mortality	5	0.16 %
XXI Factors influencing health status and contact with health services	89	2.81 %

4. Pilot data analysis

4.1. Rationale for using the Rasch technique for psychometric analysis

The rationale for using the Rasch analysis technique (Rasch 1960) to transform the WHODAS raw total score into a scale with interval scale properties is that, in principle, purely ordinally scaled values derived from the WHODAS raw total score do not allow us to calculate sums, averages or variances. An ordinal to interval transformation is essential in order to make the information collected by the questionnaire usable for measurement, comparisons, and longitudinal analysis. Consequently, a psychometric analysis using the Rasch model was performed on the entire sample of N = 3,118 applicants included in the pilot data collection.

Rasch analysis is a recognized modern test theory statistical method from the field of probabilistic measurement, first introduced in the 1960s by the Danish mathematician George Rasch. If data fits the Rasch model, the raw score of a scale can be considered fit for measurement and interval-scaled, which is a precondition for true measurement (rather than mere ordinal ranking). To analyze the WHODAS items, we used a polytomous version of the Rasch model called the Partial Credit Model (PCM) (Masters 1982).

A Rasch analysis makes it possible to test the essential preconditions that items of a questionnaire must fulfil to be used for measurement. (Bond and Fox 2001; Tennant and Conaghan 2007). These measurement preconditions, or assumptions are: (1) the targeting of the scale; (2) the reliability of the questionnaire; (3) the ordering of the items' response options; (4) the absence of strong

associations between items, called, local item independence (LID); (5) the fit of the items to the Rasch model; (6) the absence of effects of personal factors such as gender and age on item responses, called differential item functioning (DIF); and (7) the unidimensionality of the questionnaire. If these measurement assumptions can be met, a questionnaire is judged to be psychometrically sound and has the required interval-scaled total scores that are preconditions for true measurement.

For a questionnaire to truly measure degrees of functioning, the level of difficulty of the items in the questionnaire must match the population's actual level of ability, i.e., the questionnaire must not measure only high degrees or only low degrees of difficulty. Statistically, good targeting (assumption 1) means that the mean item difficulty level and mean person ability level approximate 0 and that the item difficulties match the ability of the population as a whole.

A Person Separation Index (PSI) above 0.8 is the indication of a good (assumption 2) reliability of the scale, and values above 0.9 indicate a very good reliability. The PSI indicates how well the scale can discriminate between levels of functioning in the population. The Cronbach α, which is typically also reported, is a measure of the data's internal consistency, i.e., how well the items work together to describe one construct of interest (Nunnally and Bernstein 1994), in this case the construct of functioning.

In the presence of disordered response options (assumption 3) – e.g., response 2 is higher than response 3 – an analysis of response probability curves makes it possible to determine which response options cause a disordered response problem and decide on which strategy to use in order to collapse the scale, i.e., to aggregate adjacent response options. For example, if an item's response options 2 and 1 appear reversed and indicate that an expected increase of difficulty is not observed in the data, the item responses can be recoded so that these options represent only one level of response.

Local item dependency (assumption 4) is a problem that can occur when items are redundant and measure approximately the same aspect of a construct. The most widely reported statistic for the item dependencies is the Q3 matrix, which is just another name for the Rasch residual's correlation matrix (Yen 1984). Marais (2013) recommends considering LID relative to the residual correlations' average because the residual correlation's magnitude depends on the number of items. Christensen, Makransky, and Horton (2017) formalized this, illustrating that if the largest Q3 value is more than 0.2 above the average, this indicates an anomaly that affects the validity of the construct. One way to address local item dependency without deleting questionnaire items (and thereby limiting the information collected by the questionnaire) is to aggregate or sum up the correlated items into so-called testlets (Yen 1993). In item testlets, the ordering of the thresholds is not evident.

With good item fit (assumption 5), the Infit and Outfit values are below 1.2 (R. M. Smith, Schumacker, and Bush 1998). The Outfit statistic is a more outlier-sensitive alternative to the Infit statistic, meaning that the Outfit statistic can sometimes indicate misfit, while the Infit does not.

Ideally, items of a questionnaire should not favor subgroups of the population sample. The analysis of (assumption 6) DIF using a variance analysis (ANOVA) makes it possible to flag exogenous variables (or DIF variables), which can cause a lack of invariance of the item difficulty estimates (Holland and Wainger 1993). It is worth noting that a DIF analysis does not necessarily indicate a measurement or metric bias but can also represent subgroups with an unequal ability (Boone 2016). A two-way ANOVA is used to test for uniform DIF variables as well as a non-uniform DIF (a DIF variable x score level). The questionnaire was tested for DIF by gender and age groups, which is sufficient to show the validity of the construct.

Finally, a questionnaire should measure only one construct (assumption 7). If the questionnaire presents several separate dimensions (e.g., the measure of the difficulty in the Getting Around domain

cannot be compared to the measure of difficulty of the Self Care domain) the validity of a summary total score is compromised. A principal component analysis of the residuals determines the questionnaire's degree of unidimensionality (E. V. Smith 2002). Typically, a first eigenvalue < 1.8 is deemed indicative of unidimensionality. Based on simulation analyses, R. M. Smith, and Miao (1994) suggest considering the second component's size, with values below 1.4 as indicative for unidimensionality.

The metric analyses were performed with the software R (Team 2016), specifically, the package *mirt* for the Rasch analysis (Mair, Hatzinger, and Maier 2019) and *iarm* for the DIF analysis.

4.2 Bulgaria pilot: Metric properties of WHODAS

As noted, in Bulgaria, the 36-item version of WHODAS was used. In the collected data, WHODAS items D5.5 to D5.8, which were responded to only by persons working or in education constituted more than 65 percent of all missing values in the pilot. These items were excluded from the analysis (Table 1.1). The WHODAS Manual (Ustün 2009)[12] indicates that the 32-item score is highly comparable to the 36-item score, so excluding these items does not have any measurement impact. The psychometric analysis therefore included the remaining 32 items.

The work items D5.5 to D5.8 – which as noted resulted in > 65.0 percent of missing values – represent the responses of persons who were on average significantly older (58.63 years of age, SD = 14.67) compared to those who responded (51.27 years, SD = 9.73). These individuals were either retired (39.8 percent) or unemployed for health reasons (53.3 percent). The psychometric analysis did not include these items. Two other items had higher percentages of missing values (> 5.0 percent): *D3.4 Staying by yourself for a few days* (5.52 percent), and *D4.5 Sexual activities* (28.03 percent). While the Rasch model can still handle percentages below 15.0 percent without introducing detrimental bias in the estimates (Fellinghauer, Prodinger, and Tennant 2018), percentages of missing values above 50.0 percent are not acceptable.[13]

Based on the WHODAS Manual, the simple approach to imputing missing values is to use the mean of a person's score, but only if 1 or 2 items are with missing values; if more than two items are missing this method should not be used. Notwithstanding the work-related items' missing values, the data collection in Bulgaria had a very good response rate, with only N = 46 (1.48 percent) of the WHODAS survey participants having more than two missing values. A total of N = 2,073, i.e., 66.48 percent of the pilot participant did not have any unanswered items on the 32 items used for the psychometric analysis. A statistical imputation approach including the socio-demographic information could be used for the missing values in the WHODAS items, but given that small number of missing values, imputation was not thought to be necessary.

Figure 2 shows the distribution of the WHODAS total score without or with data imputation. The red lines show the 25.0 percent distribution quantile, the blue dotted line the median, and the green line the 75.0 percent distribution quantile.

[12] Ustun, Tevfik Bedirhan, Kostanjesek, N, Chatterji, S, Rehm, J & World Health Organization. (2010). Measuring health and disability: manual for WHO Disability Assessment Schedule (WHODAS 2.0) / edited by T.B. Üstün, N. Kostanjsek, S. Chatterji, J. Rehm. World Health Organization. https://apps.who.int/iris/handle/10665/43974/

[13] It should be noted that these same items also recorded higher proportions of missing values in the other World Bank's WHODAS pilots recently conducted in Lithuania and Latvia.

Figure 2: Bulgaria WHODAS Pilot Raw Score Distributions

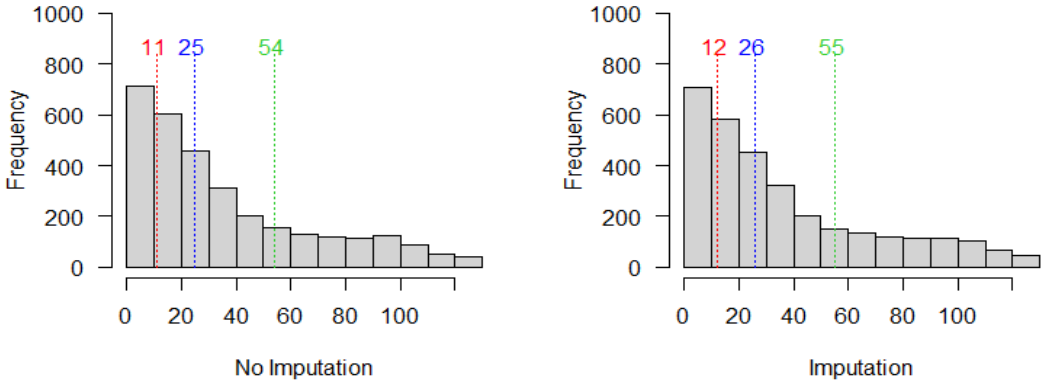


Table 7 shows the descriptive statistics for the 36 WHODAS items, including the number and percentage of missing values. A higher percentage of the applicants indicated severe or extreme problems in D2.5 – *Walking a long distance such as a kilometer?* (40.51 percent), D6.6 – *How much has your health been a drain on the financial resources for you or your family?* (37.3 percent). About 30.0 percent of the sample reported severe to extreme problems in D2.1 – *Standing for long periods such as 30 minutes?* (28.15 percent), D5.3 – *Getting all the household work done that you needed to do?* (30.63 percent), D5.4 -- *Getting your household work done as quickly as needed?* (29.76 percent), D6.4 – *How much time did you spend on your health condition or its consequences?* (31.68 percent), D6.5 – *How much have you been emotionally affected by your health condition?* (33.32 percent), D6.7 – *How much of a problem did your family have because of your health problems?* (29.63 percent), and D6.8 – *How much of a problem did you have in doing things by yourself for relaxation or pleasure?* (30.34 percent).

Table 7: Frequencies and percentages of WHODAS responses

Item	No	Mild	Moderate	Severe	Extreme, cannot do	Missing
D1.1	2044 (65.55%)	458 (14.69%)	328 (10.52%)	170 (5.45%)	118 (3.78%)	0 (0%)
D1.2	1824 (58.5%)	573 (18.38%)	367 (11.77%)	186 (5.97%)	168 (5.39%)	0 (0%)
D1.3	1836 (58.88%)	478 (15.33%)	353 (11.32%)	219 (7.02%)	232 (7.44%)	0 (0%)
D1.4	1850 (59.33%)	455 (14.59%)	318 (10.2%)	230 (7.38%)	265 (8.5%)	0 (0%)
D1.5	2479 (79.51%)	267 (8.56%)	173 (5.55%)	130 (4.17%)	69 (2.21%)	0 (0%)
D1.6	2365 (75.85%)	315 (10.1%)	201 (6.45%)	121 (3.88%)	116 (3.72%)	0 (0%)
D2.1	1001 (32.1%)	607 (19.47%)	632 (20.27%)	427 (13.69%)	451 (14.46%)	0 (0%)
D2.2	1277 (40.96%)	723 (23.19%)	539 (17.29%)	300 (9.62%)	279 (8.95%)	0 (0%)
D2.3	1862 (59.72%)	476 (15.27%)	307 (9.85%)	217 (6.96%)	256 (8.21%)	0 (0%)
D2.4	1399 (44.87%)	507 (16.26%)	455 (14.59%)	304 (9.75%)	453 (14.53%)	0 (0%)
D2.5	712 (22.84%)	520 (16.68%)	560 (17.96%)	487 (15.62%)	776 (24.89%)	63 (2.02%)
D3.1	2037 (65.33%)	348 (11.16%)	234 (7.5%)	186 (5.97%)	313 (10.04%)	0 (0%)
D3.2	2137 (68.54%)	327 (10.49%)	233 (7.47%)	184 (5.9%)	237 (7.6%)	0 (0%)
D3.3	2521 (80.85%)	194 (6.22%)	182 (5.84%)	113 (3.62%)	108 (3.46%)	0 (0%)
D3.4	1759 (56.41%)	287 (9.2%)	203 (6.51%)	175 (5.61%)	522 (16.74%)	172 (5.52%)
D4.1	2105 (67.51%)	393 (12.6%)	205 (6.57%)	174 (5.58%)	241 (7.73%)	0 (0%)
D4.2	2318 (74.34%)	268 (8.6%)	180 (5.77%)	171 (5.48%)	181 (5.81%)	0 (0%)
D4.3	2552 (81.85%)	243 (7.79%)	155 (4.97%)	97 (3.11%)	71 (2.28%)	0 (0%)
D4.4	1926 (61.77%)	416 (13.34%)	235 (7.54%)	192 (6.16%)	294 (9.43%)	55 (1.76%)
D4.5	1169 (37.49%)	313 (10.04%)	246 (7.89%)	172 (5.52%)	344 (11.03%)	874 (28.03%)
D5.1	1226 (39.32%)	689 (22.1%)	479 (15.36%)	230 (7.38%)	494 (15.84%)	0 (0%)
D5.2	1167 (37.43%)	681 (21.84%)	505 (16.2%)	246 (7.89%)	519 (16.65%)	0 (0%)
D5.3	814 (26.11%)	751 (24.09%)	598 (19.18%)	337 (10.81%)	618 (19.82%)	0 (0%)
D5.4	865 (27.74%)	750 (24.05%)	575 (18.44%)	333 (10.68%)	595 (19.08%)	0 (0%)
D5.5	687 (22.03%)	185 (5.93%)	95 (3.05%)	35 (1.12%)	34 (1.09%)	2082 (66.77%)
D5.6	702 (22.51%)	164 (5.26%)	111 (3.56%)	28 (0.9%)	31 (0.99%)	2082 (66.77%)
D5.7	610 (19.56%)	232 (7.44%)	113 (3.62%)	46 (1.48%)	35 (1.12%)	2082 (66.77%)
D5.8	627 (20.11%)	232 (7.44%)	97 (3.11%)	48 (1.54%)	32 (1.03%)	2082 (66.77%)
D6.1	1312 (42.08%)	525 (16.84%)	392 (12.57%)	309 (9.91%)	465 (14.91%)	115 (3.69%)
D6.2	1321 (42.37%)	676 (21.68%)	441 (14.14%)	421 (13.5%)	259 (8.31%)	0 (0%)
D6.3	1645 (52.76%)	556 (17.83%)	433 (13.89%)	313 (10.04%)	171 (5.48%)	0 (0%)
D6.4	589 (18.89%)	756 (24.25%)	785 (25.18%)	740 (23.73%)	248 (7.95%)	0 (0%)
D6.5	548 (17.58%)	756 (24.25%)	775 (24.86%)	753 (24.15%)	286 (9.17%)	0 (0%)
D6.6	641 (20.56%)	565 (18.12%)	749 (24.02%)	937 (30.05%)	226 (7.25%)	0 (0%)
D6.7	764 (24.5%)	738 (23.67%)	692 (22.19%)	721 (23.12%)	203 (6.51%)	0 (0%)
D6.8	1112 (35.66%)	578 (18.54%)	482 (15.46%)	366 (11.74%)	580 (18.6%)	0 (0%)

Table 8 and Table 9 provide the essential fit statistics for the WHODAS items at the start and after the metric adjustments based on the outcomes of the Rasch-based analysis with the Partial Credit Model. The whole scale showed multidimensionality with a strong tendency of the items to load by WHODAS

domains. However, a few items cross-loaded, and a few items were free of dependencies (Figure 2). The multidimensionality caused by the WHODAS item dependencies was adjusted first by aggregating the items into testlets based on the WHODAS domain structure. A significant local dependency remained between D2 - *Getting around* and D3 - *Self-care*, as well as D1 - *Understanding and Communicating* and D4 - *Getting along with people*. These domains were also aggregated. This adjustment strategy worked well, with good reliability. A few items presented higher infit and outfit values, however the testlets that adjust for the item dependencies presented good fit. The analysis was undertaken with non-imputed data; however, the model fit statistics are also shown for imputed data to indicate that imputation would not alter the model targeting and fit. The missing value imputation was performed with *MissForest*, a robust multiple imputation method for mixed-type data (Stekhoven and Buhlmann 2012). Table 10 shows the starting approach's reliability statistics after aggregating the items for all dependencies among domains, without imputation, and with data imputation.

Table 8: WHODAS item difficulties, fit, local item dependencies, and differential item functioning at the start

WHODAS Item Number	Outfit[1]	Infit[1]	Item Difficulty	Disordered Thresholds	LID[3]	DIF[4]
D1.1	1.12	1.21	2.03	x	D1.2, D1.3, D1.4, D1.5, D1.6, D4.1, D4.2, D4.3	Age
D1.2	0.94	1.05	1.74		D1.1, D1.3, D1.4, D1.5, D1.6, D4.1, D4.2, D4.3, D4.4	Age
D1.3	0.82	0.99	1.53	x	D1.1, D1.2, D1.4, D1.5, D1.6, D4.1, D4.2, D4.3, D4.4	Age
D1.4	0.75	0.92	1.46	x	D1.1, D1.2, D1.3, D1.5, D1.6, D4.1, D4.2, D4.4	Age
D1.5	0.85	1.07	2.6	x	D1.1, D1.2, D1.3, D1.4, D1.6, D4.1, D4.2, D4.3, D4.4	Age
D1.6	0.80	1.13	2.29	x	D1.1, D1.2, D1.3, D1.4, D1.5, D4.1, D4.2, D4.3, D4.4	Age
D2.1	1.24	1.29	0.55		D2.2, D2.3, D2.4, D2.5	Age
D2.2	1.05	1.19	1.09		D2.1, D2.3, D2.4, D2.5, D3.2	Age
D2.3	0.62	0.90	1.5	x	D2.1, D2.2, D2.4, D2.5, D3.1, D3.2	Age
D2.4	0.67	0.81	0.84	x	D2.1, D2.2, D2.3, D2.5	Age
D2.5	1.28	1.23	-0.08		D2.1, D2.2, D2.3, D2.4	Age
D3.1	0.51	0.71	1.49	x	D2.3, D3.2, D3.3, D5.1	
D3.2	0.52	0.71	1.71	x	D2.2, D2.3, D3.1, D3.3	
D3.3	0.43	0.87	2.41	x	D3.1, D3.2	
D3.4	0.58	0.71	0.99	x		
D4.1	0.99	0.97	1.71	x	D1.1, D1.2, D1.3, D1.4, D1.5, D1.6, D4.2, D4.3, D4.4	Age
D4.2	0.65	0.90	1.98	x	D1.1, D1.2, D1.3, D1.4, D1.5, D1.6, D4.1, D4.3, D4.4	Age
D4.3	0.40	0.86	2.68	x	D1.1, D1.2, D1.3, D1.5, D1.6, D4.1, D4.2, D4.4	Age
D4.4	0.78	0.89	1.48	x	D1.2, D1.3, D1.4, D1.5, D1.6, D4.1, D4.2, D4.3	Age
D4.5	1.68	1.36	0.62	x		Age
D5.1	0.50	0.55	0.74	x	D3.1, D5.2, D5.3, D5.4	
D5.2	0.46	0.52	0.66	x	D5.1, D5.3, D5.4	
D5.3	0.52	0.55	0.28	x	D5.1, D5.2, D5.4	
D5.4	0.49	0.53	0.34	x	D5.1, D5.2, D5.3	
D6.1	0.64	0.73	0.69		no	
D6.2	0.75	0.83	0.79		D6.3	
D6.3	1.22	1.28	1.25		D6.2	Age
D6.4	1.16	1.17	1.17		D6.5, D6.6	
D6.5	0.97	0.99	0.98		D6.4, D6.6, D6.7	Age, Gender
D6.6	1.21	1.21	1.21		D6.4, D6.5, D6.7	Age, Gender
D6.7	0.87	0.92	0.89		D6.5, D6.6	Age
D6.8	0.66	0.76	0.71		no	

1 Infit and Outfit expected below 1.2 for the absence of underfit
2 In testlets, i.e., aggregated locally dependent items, the ordering of thresholds is not expected anymore
3 Local item dependency (LID) significant if LID > mean residual correlation + 0.2
4 Differential item functioning (DIF)

Table 9: WHODAS Item Difficulties, fit, local item dependencies, and differential item functioning after adjustments

Label	WHODAS Item Number	Outfit[1]	Infit[1]	Item Difficulty	Disordered Thresholds	LID[3]	DIF[4]
Testlet 1	D1.1-D1.6 & D4.1-D4.5	0.97	1.17		n.a.[2]	no	Age
Testlet 2	D2.1-D2.5 & D3.1-D3.4	0.75	0.89	0.36	n.a.[2]	no	Age
Testlet 3	D5.1-D5.4	0.49	0.52	0.16	n.a.[2]	no	Age
Testlet 4	D6.1-D6.8	0.61	0.64	0.20	n.a.[2]	no	Age

1 Infit and Outfit expected below 1.2 for the absence of underfit
2 In testlets, i.e., aggregated locally dependent items, the ordering of thresholds is not expected anymore
3 Local item dependency (LID) significant with r > 0.2
4 Differential item functioning (DIF)

Table 10: Targeting and Reliability of WHODAS items

	Start		**1) Domain based item aggregation**		**2) Domain based item aggregation with imputed data**	
	Targeting		**Targeting**		**Targeting**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Difficulty	1.2	1.06	0.33	0.62	0.33	0.57
Ability	0	1.65	-0.01	0.54	-0.01	0.56
	PSI	**Alpha**	**PSI**	**Alpha**	**PSI**	**Alpha**
Reliability	0.95	0.98	0.84	0.91	0.85	0.91

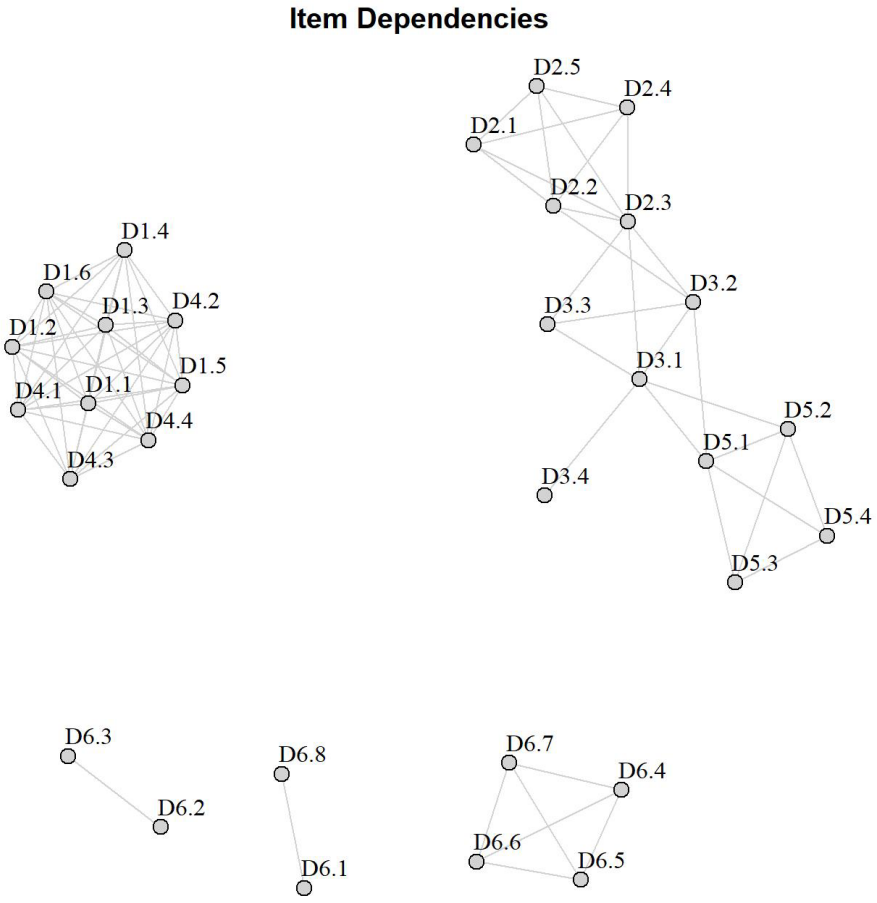
Next, we report more details on the *psychometric analysis* of the WHODAS with 32 items at the start and the final approach (without imputation), deemed the most efficient and resulting in best metric properties. The final model retained the 32 items, with domains aggregated based on a conceptual, domain-based approach, preserving the assessment tool's structure. Table 8-10 above, present the detailed Rasch statistics. Table 10 includes the targeting and reliability with imputed data.

1. The targeting of the scale (Table 10) improved with adjustments, i.e., item difficulties becoming more centered on the mean functioning level.
2. Item dependencies (Figure 3) in the analysis with the 32 WHODAS items inflated the reliability estimates (PSI = 0.95, Cronbach α = 0.95). After adjustment for the local item dependencies, the reliability dropped to PSI = 0.84, which is still a good level of reliability, indicating that the metric can discriminate well among levels of functioning (Table 10).
3. Twelve items only presented perfectly ordered difficulty threshold. Disordering consisted of a reversing of either the 2 lower categories (0 = None & 1 = Mild) or the 2 upper (3 = Severe & 4 = Extreme) categories (Figure 4). A perfect ordering could be obtained on the entire 32 items scale by reducing the number of response options through collapsing: [0 & 1] = 0, [2 & 3] = 1, and [4 = Extreme] = 2 (Figure 5). However, the collapsing of the response options alone would not solve the multidimensionality and item dependencies so that other adjustment measures would still be required. Figure 5 shows the person item map for the approach with collapsed response options. Ordering of thresholds is not expected anymore with testlets.
4. The residual dependencies analysis indicated strong local dependencies among the 32 items of the WHODAS (see Figure 4), with a tendency of questionnaire items from the same domain to

associate. The cut-off for LID using the mean LID + 0.2 as cut-off (Christensen, Makransky, and Horton 2017) was r = 0.17. To solve these dependencies, the items were aggregated, taking into account the chapter structure, i.e., D1 - *Understanding and Communicating*, D2 - *Getting around*, D3 - *Self-Care*, D4 - *Getting along with people*, D5 - *Life Activities*, D6 - *Participation in Society*. A significant residual correlation above r > 0.018 was still found between D1 - *Understanding and Communicating*) and D4 - *Getting along with people* (r = 0.34) and D2 - *Getting around* - and D3 *Self-care* (r = 0.13) which were also aggregated.

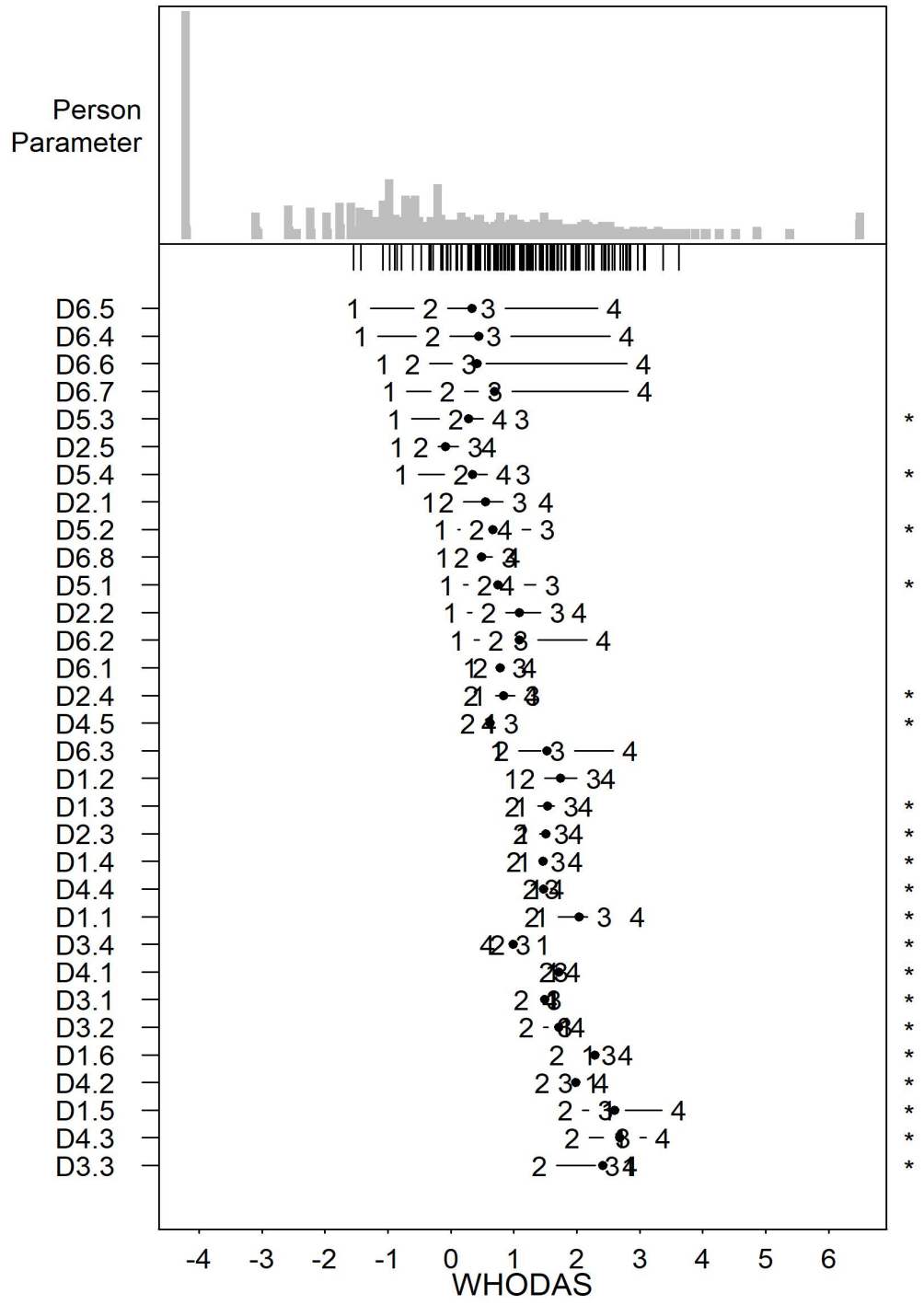
5. With Infit and Outfit expected below 1.2, a few misfitting items can be observed at the start, most being very close to the cut-off of 1.2. Only the item D4.5 *Sexual activities* (MSQ_{Infit} = 1.36, MSQ_{Outfit} = 1.68) stronger departs from good fit (Table 8). After aggregation of the dependent items, the testlets did not show any underfit based on the Infit and the Outfit statistic (Table 9).
6. DIF was tested for gender and age. In the final model, all testlets presented some DIF for the age groups. It can be expected that the levels of functioning and disability are impacted by the age. The residuals did not present any pattern indicating DIF for gender (Table 9).

Figure 3: Local item dependencies before the creation of testlets



Finally, the principal component analysis indicated clustering of the items by domains resulting in multidimensionality, with a 1^{st} eigenvalue of 6.88 and a 2^{nd} eigenvalue of 2.99. After adjustments, i.e., aggregation of items by WHODAS domains, the 1^{st} eigenvalue dropped to 1.86 and the 2^{nd} eigenvalue to 1.21 and supported unidimensionality.

Figure 4: Person Item Map for the WHODAS items without Rasch-based adjustment



**Indicate disordered thresholds*

Figure 5: Person Item Map after collapsing of the response options into three categories

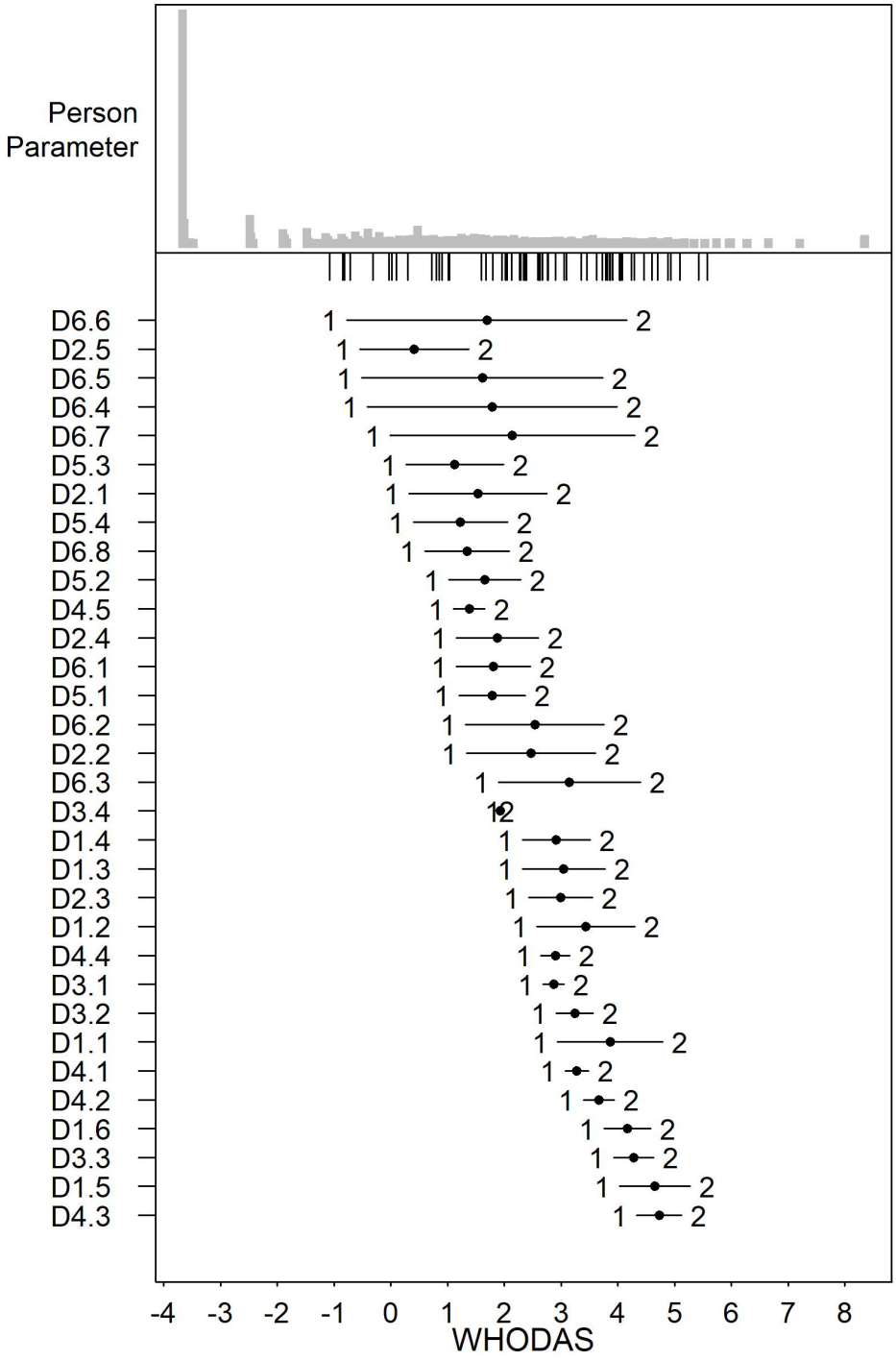
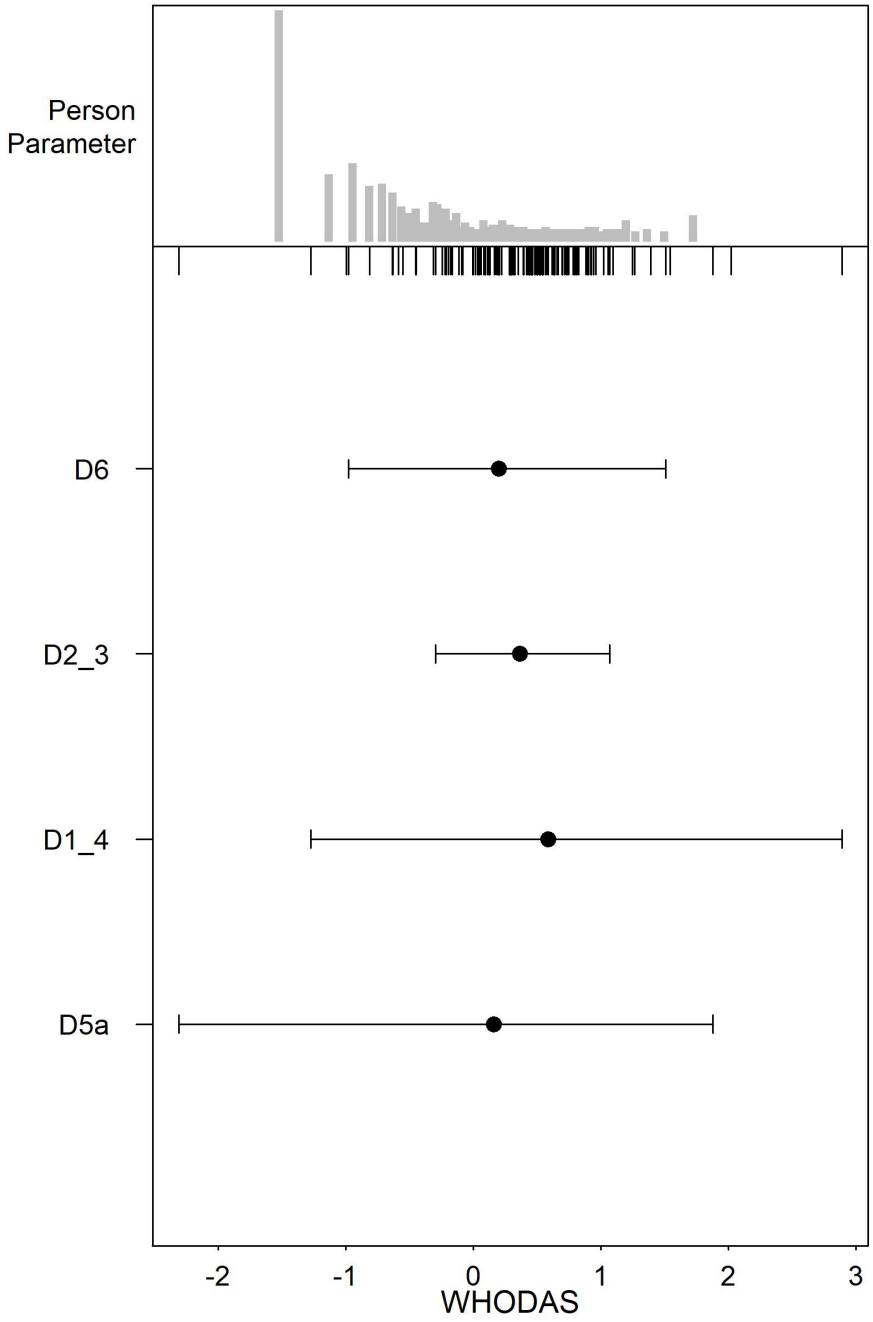


Figure 6: Person Item Map for the WHODAS after the creation of testlets



Finally, Table 11 gives the final result, namely, the WHODAS score transformations (this required logit-scaled Rasch ability estimates that made it possible to recode scores from the 32 WHODAS items into a psychometrically sound 0-100 interval-scaled metric).

Table 11: Transformation Table

WHODAS 2.0 Score	Rasch Logit	0-100 Score	WHODAS 2.0* Score	Rasch* Logit	0-100* Score
32	-1.52	0	96	0.45	62
33	-1.13	12	97	0.45	63
34	-0.94	18	98	0.46	63
35	-0.82	22	99	0.47	63
36	-0.72	26	100	0.47	63
37	-0.64	28	101	0.48	63
38	-0.56	30	102	0.49	64
39	-0.5	32	103	0.49	64
40	-0.45	34	104	0.5	64
41	-0.4	36	105	0.51	64
42	-0.36	37	106	0.51	64
43	-0.32	38	107	0.52	65
44	-0.28	39	108	0.53	65
45	-0.25	40	109	0.53	65
46	-0.22	41	110	0.54	65
47	-0.19	42	111	0.55	66
48	-0.16	43	112	0.55	66
49	-0.13	44	113	0.56	66
50	-0.11	45	114	0.57	66
51	-0.09	46	115	0.57	66
52	-0.06	46	116	0.58	67
53	-0.04	47	117	0.59	67
54	-0.02	48	118	0.59	67
55	0	48	119	0.6	67
56	0.01	49	120	0.61	67
57	0.03	49	121	0.61	68
58	0.05	50	122	0.62	68
59	0.06	50	123	0.63	68
60	0.08	51	124	0.64	68
61	0.1	51	125	0.64	69
62	0.11	52	126	0.65	69
63	0.13	52	127	0.66	69
64	0.14	53	128	0.67	69

65	0.15	53	129	0.68	70
66	0.17	53	130	0.68	70
67	0.18	54	131	0.69	70
68	0.19	54	132	0.7	70
69	0.2	55	133	0.71	71
70	0.22	55	134	0.72	71
71	0.23	55	135	0.73	71
72	0.24	56	136	0.74	72
73	0.25	56	137	0.75	72
74	0.26	56	138	0.76	72
75	0.27	57	139	0.77	73
76	0.28	57	140	0.78	73
77	0.29	57	141	0.79	73
78	0.3	58	142	0.81	74
79	0.31	58	143	0.82	74
80	0.32	58	144	0.84	75
81	0.33	59	145	0.85	75
82	0.34	59	146	0.87	76
83	0.34	59	147	0.89	76
84	0.35	59	148	0.91	77
85	0.36	60	149	0.93	78
86	0.37	60	150	0.95	78
87	0.38	60	151	0.98	79
88	0.39	60	152	1.01	80
89	0.39	61	153	1.05	81
90	0.4	61	154	1.09	83
91	0.41	61	155	1.14	84
92	0.42	61	156	1.2	86
93	0.42	62	157	1.27	88
94	0.43	62	158	1.36	91
95	0.44	62	159	1.5	96
96	0.45	62	160	1.64	100

4.3 A summary and conclusions about WHODAS psychometric properties in the Bulgarian pilot

The WHODAS with 32 items, leaving out the work and education items, required only a few adjustments. Aggregation of the items by chapter domains solved the multidimensionality caused by the domain-based item cluster and delivered an unbiased reliability estimate. This adjustment approach preserved the original conceptual form of the instrument, i.e., it validly captures the construct of disability. The instrument is sound and delivers a reliable interval scaled score when items are considered by domains. The DIF found for age groups in the testlets indicate that with increasing age the levels of functioning decrease, resulting in higher levels of disability. A conversion table (Table 11) transforms the raw scores on the 32 items of the WHODAS into 0-100-point measure of disability.

One must note that for WHODAS the disordering of the response options was prominent. This can be caused by the dependencies among items and missing values in the data. Another plausible explanation is that the applicants when answering questions have difficulties differentiating 'none' from 'mild' or 'moderate' from 'severe' functioning problems. Ideally, a well-trained interviewer would ask probing questions and support the respondent to describe her or his performance in the best possible way, so that interviewer can then choose the right response option.

The seven essential statistical tests described above show that the data collected with WHODAS, under the Rasch analysis, display robust psychometric properties of validity and reliability. *With a few adjustments, the scale is unidimensional and free of item dependencies with good targeting and with good reliability. Aggregating the items by domains solves observed local item dependencies and produces a unidimensional assessment metric. The domain-based testlets fit well, and a transformation table is obtained that translates the observed sum scores into an interval-scaled metric.*

It is important to keep in mind that the WHO developed WHODAS explicitly to statistically capture the construct of functioning from the perspective of performance – namely the real experience of performing activities by a person with an underlying health problem in their everyday life. There is an abundance of evidence from the scientific literature – supported by the results of this pilot – that WHODAS is a psychometrically sound instrument that reliably and validly collects information about levels of disability.

Therefore, we can confidently conclude that information collected with the WHODAS is robust, viable and relevant and that it validly represents the construct of disability as understood in ICF and the Convention on the Rights of Rights of Persons with Disabilities (CRPD). It can, thus, surely be included into the disability status assessment in Bulgaria to: (i) significantly strengthen the method of assessment currently in use (mostly based on impairments); (ii) bring it closer to the ICF and CRPD understanding of disability; and (iii) harmonize the approach to assessment with the approach used in the individual needs assessment.

4.4 Comparing WHODAS and certified disability degree data

4.4.1 Comparing WHODAS and certified disability scores

One of the objectives of our analysis of the WHODAS data collected in Bulgaria is to show that the inclusion of functioning into the current medically based disability assessment method will significantly improve its capacity to assess the experience of disability more accurately.

Our data set included WHODAS data, and a certified disability (percentage) as determined by TMECs/NMEC, as well as associated ICD codes for each participant in the WHODAS survey. This allowed us to compare the two sets of data. Below we present relevant descriptive statistics regarding the medically determined disability severity ranking by the Bulgarian disability assessment system and

the Rasch-based WHODAS scores for the same individuals from the pilot, taking into account socio-demographic characteristics and main ICD diagnoses.

As mentioned, the current disability assessment method is medically based and uses an instrument (a Baremic type instrument) that matches diseases and associated impairments with percentages of disability (as compared to a healthy person). Procedurally, the assessment is based on medical documentation, justifying the degree of impairments in diseased or injured body part or structure, detailed relevant clinical history, in-depth clinical examination, and, in some cases, targeted laboratory and examinations performed by TMEC/NMEC.[14] The current Bulgarian disability assessment system identifies disability levels as percentages – with values < 50 percent designating mild/no disability, 50-70 percent moderate disability, 70-90 percent severe disability, and > 90 percent very severe disability. In what follows we will call these here 'disability severity ranking groups'.

Table 12 presents the WHODAS functioning score by current Bulgarian disability severity ranking groups. The increase in the WHODAS disability score is weakly matched by the medical assessment severity ranking for moderate and severe disability. The mean WHODAS score in the group of persons with moderate disability severity is 39.6 (SD = 14.3) and the severe disability is 41.8 (SD = 14.9). The difference of only 5.0 percent indicates that moderate and severe disability are not differentiated well in the current disability assessment system, suggesting low reliability and precision. The match is stronger in the case of very severe disability where the mean WHODAS Rasch score is noticeably higher, 57.3 (SD = 14.3).

Table 12: WHODAS scores distribution for the pilot sample and by disability severity category: mean, standard deviation and quartiles

	Mean	SD	25%	50%	75%
Pilot	46.7	16.6	37.2	48	58.4
Moderate	39.6	14.3	32.7	42.1	49.5
Severe	41.8	14.9	34.6	45	52.5
Very Severe	57.3	14.3	49.4	60	66.3

Figure 7 shows the density lines per disability severity ranking group and makes it clear that there is a gradient of disability that can be observed across the medical disability severity ranking groups. While the WHODAS scores for very severe functioning restrictions stand out (red line), the difference between severe and moderate disability severity (the yellow and orange lines) is less obvious with a closer location to each other. The observed gradient of disability seen in Figure 7 is confirmed by the ANOVA-test of the WHODAS-scores by disability severity ranking group, which is significant ($F_{(df)} = 487.3_{(2)}$; $p - value < 0.001$. Table 13). The Tuckey Test confirms that the discrimination is weaker between moderate and severe disability severity ranking, as also shown in Figure 7, where the density lines of these two groups are closer (Table 14).

The density lines in Figure 7 also suggest the presence of false positives (high disability percentage and low WHODAS score) and false negatives (lower disability percentage and high WHODAS score). A more accurate assessment would show the very severe WHODAS density line sloping more to the right-hand side of the Figure 7: the line would be close to 0 up until the score of 45, then sharply rising around the score of 50. The opposite should be the case for the moderate disability, which should be

[14] See Chapter 3 of the above-mentioned World Bank Report.

located mostly to the left-hand side of the Figure 7. The medical information may misrepresent the true extent of individual disability as experienced in daily life.

Figure 7: WHODAS score density line by disability severity

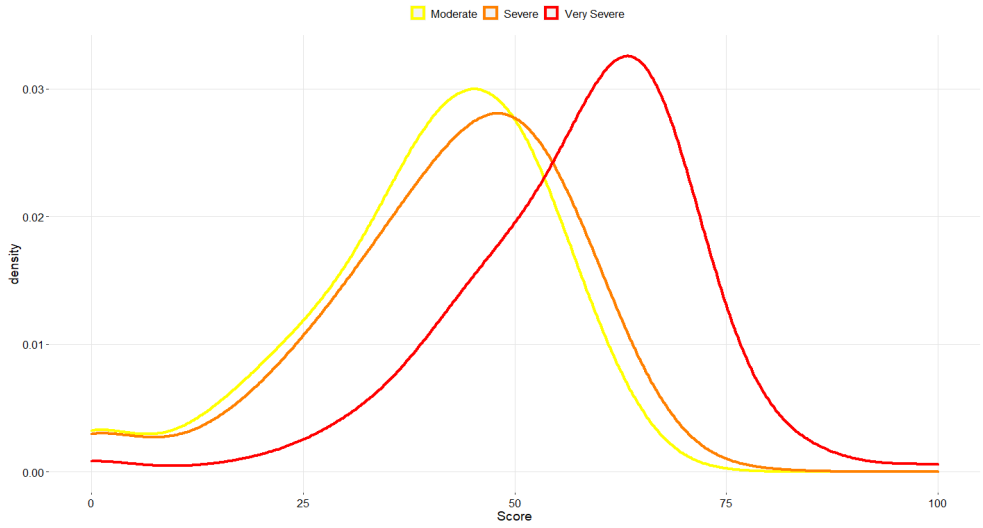


Table 13: Analysis of variance of the WHODAS scores by disability severity

	DF	Sum of squares	Mean squares	F-value	P-value
Disability severity	2	203769	101884	487.3	< 0.001
Residuals	3115	651316	209		

Table 14: Tukey Test for the WHODAS scores by disability severity

	Difference	Lower CI	Upper CI	P-values
Severe-Moderate	2.178	0.649	3.706	0.002
Very Severe-Moderate	17.702	16.258	19.145	< 0.001
Very Severe-Severe	15.524	14.016	17.032	< 0.001

Sample characteristics vary significantly with increasing disability percentage. Table 15 shows the socio-demographic characteristics of the pilot sample disaggregated by disability severity ranking groups, including p-value for a significance test comparing each socio-demographic variable by disability severity ranking groups. The group with more severe disability is older, with a mean age of 60.91 (SD = 15.79), more likely to be retired (44.2 percent) or unemployed for health reasons (39.3 percent).

Table 15: Pilot sample descriptive statistics by disability severity ranking group

	Moderate	Severe	Very Severe	P-value
Pilot Sample Size	1072	910	1136	
Gender = Male (%)	459 (42.8)	443 (48.7)	547 (48.2)	0.012
Age - mean (SD)	54.29 (10.22)	52.52 (12.73)	60.91 (15.79)	*<0.001*
Years of Education - mean (SD)	11.92 (3.23)	12.19 (3.22)	11.14 (3.56)	*<0.001*
Marital Status (%)				*<0.001*
Never married	158 (14.7)	213 (23.4)	216 (19.0)	
Currently married	612 (57.1)	437 (48.0)	531 (46.7)	
Separated	18 (1.7)	15 (1.6)	11 (1.0)	
Divorced	133 (12.4)	118 (13.0)	121 (10.7)	
Widowed	95 (8.9)	70 (7.7)	223 (19.6)	
Cohabiting	56 (5.2)	57 (6.3)	34 (3.0)	
Living Condition (%)				0.032
Independent in the community	1070 (99.8)	903 (99.2)	1122 (98.8)	
Assisted living	2 (0.2)	3 (0.3)	10 (0.9)	
Hospitalized	0 (0.0)	4 (0.4)	4 (0.4)	
Work Status (%)				n.a.
Paid work	470 (44.5)	332 (37.1)	143 (12.8)	
Self-employed	44 (4.2)	24 (2.7)	11 (1.0)	
Non-paid work	0 (0.0)	0 (0.0)	0 (0.0)	
Student	2 (0.2)	3 (0.3)	7 (0.6)	
Keeping house	5 (0.5)	3 (0.3)	2 (0.2)	
Retired	145 (13.7)	171 (19.1)	495 (44.2)	
Unemployed (health reasons)	330 (31.2)	318 (35.5)	441 (39.3)	
Unemployed (other reasons)	61 (5.8)	43 (4.8)	20 (1.8)	
Other	0 (0.0)	1 (0.1)	2 (0.2)	

Table 16 shows, for the sample and disaggregated by disability severity ranking group, the number and percentage of reported primary ICD condition chapters, as well as the corresponding mean and standard deviation of the WHODAS score. The most frequently reported conditions are *II Neoplasms* (N = 758, 24.31 percent), *IX Diseases of the circulatory system* (N = 698, 22.39 percent), *IV Endocrine nutritional and metabolic diseases* (N = 245, 7.86 percent), *V Mental and behavioral disorders* (N = 342, 10.97 percent), and *XIII Diseases of the musculoskeletal system and connective tissue* (N = 255, 8.18 percent). With moderate or severe disability *IV Endocrine, nutritional, and metabolic diseases* represent more than 10.0 percent of the participants per group. In the groups of persons with severe or very severe disability percentages, *VI Diseases of the nervous system* are also reported as a main condition for more than 5.0 percent of the WHODAS survey participants.

Table 16: Frequency and Percentage of ICD chapters for the pilot sample and by disability severity ranking group as well as the mean and standard deviation (SD) of the corresponding WHODAS scores

ICD-10 Chapter	N	Mean(SD)	N Moderate	Mean(SD) Moderate	N Severe	Mean(SD) Severe	N Very Severe	Mean(SD) Very Severe
I Certain infectious and parasitic diseases	11(0.35%)	52.16 (13.68)	2(0.19%)	37.65 (18.04)	1(0.11%)	44.99	8(0.7%)	56.69 (11.39)
II Neoplasms	758(24.31%)	40.72 (15.96)	203(18.94%)	33.74 (13.49)	228(25.05%)	35.93 (15.09)	327(28.79%)	48.39 (14.65)
III Diseases of the blood	14(0.45%)	34.48 (17.85)	4(0.37%)	43.91 (7.69)	4(0.44%)	34.8 (24.28)	6(0.53%)	27.97 (17.81)
IV Endocrine nutritional and metabolic diseases	245(7.86%)	38.85 (16.23)	115(10.73%)	39.41 (13.79)	103(11.32%)	35.16 (17.61)	27(2.38%)	50.53 (14.99)
IX Diseases of the circulatory system	698(22.39%)	50.09 (15.73)	283(26.4%)	42.88 (13.32)	181(19.89%)	44.32 (13.79)	234(20.6%)	63.29 (10.69)
V Mental and behavioural disorders	342(10.97%)	56.75 (14.44)	67(6.25%)	46.65 (12.31)	123(13.52%)	50.49 (12.44)	152(13.38%)	66.26 (10.54)
VI Diseases of the nervous system	204(6.54%)	52.67 (14.98)	36(3.36%)	41.07 (12.91)	77(8.46%)	44.41 (11.14)	91(8.01%)	64.24 (9.43)
VII Diseases of the eye and adnexa	161(5.16%)	46.58 (18.5)	62(5.78%)	34.59 (18.17)	30(3.3%)	41.92 (14.73)	69(6.07%)	59.38 (10.47)
VIII Disease of the ear and mastoid process	37(1.19%)	37.05 (17.84)	26(2.43%)	32.1 (17.16)	4(0.44%)	48.84 (15.52)	7(0.62%)	48.71 (14.32)
X Diseases of the respiratory system	44(1.41%)	45.96 (14.46)	30(2.8%)	43.58 (13.86)	7(0.77%)	41.16 (11.88)	7(0.62%)	60.94 (10.56)
XI Diseases of the digestive system	53(1.7%)	45.59 (16.4)	17(1.59%)	40.14 (14.96)	17(1.87%)	44.96 (11.19)	19(1.67%)	51.03 (20.17)
XII Diseases of the skin and the subcutaneous tissue	22(0.71%)	37.86 (18.44)	18(1.68%)	33.9 (17.12)			4(0.35%)	55.64 (14.32)
XIII Diseases of the musculoskeletal system and connective tissue	255(8.18%)	46.8 (13.16)	108(10.07%)	42.42 (11.54)	83(9.12%)	43.82 (12.25)	64(5.63%)	58.06 (10.15)
XIV Diseases of the genitourinary system	57(1.83%)	46.29 (12.79)	18(1.68%)	40.2 (13.77)	8(0.88%)	40.65 (9.23)	31(2.73%)	51.28 (11.03)
XIX Injury poisoning and certain other consequences of external causes	97(3.11%)	53.21 (14.65)	35(3.26%)	42.06 (11.89)	10(1.1%)	44.97 (16.38)	52(4.58%)	62.3 (8.9)
XVII Congenital malformations deformations and chromosomal abnormalities	25(0.8%)	38.69 (17.36)	12(1.12%)	29.18 (17.32)	7(0.77%)	42.87 (13.69)	6(0.53%)	52.84 (8.9)
XVIII Symptoms signs and abnormal clinical and laboratory findings	1(0.03%)	41.21 (NA)	1(0.09%)	41.21				
XXI External causes of morbidity and mortality	5(0.16%)	52.13 (22.62)	1(0.09%)	12.06 (NA)			4(0.35%)	62.15 (3.61)
XXI Factors influencing health status and contact with health services	89(2.85%)	43.84 (14.04)	34(3.17%)	39.27 (12.99)	27(2.97%)	43.08 (11.88)	28(2.46%)	50.12 (15.24)

It can be expected that the WHODAS is more able to capture disability in persons with V *Mental and behavioral* disorders than the approach applied by the current system. The high mean WHODAS score for the individuals with a mental or behavioral disorders in the group of persons with moderate disability severity ranking is suspicious and suggests that disability in many persons with mental disorders may have been underrated. In the group of persons with a disability severity ranked as moderate the *Mental and behavioral disorders* as a primary condition have the highest mean WHODAS score (46.65, SD = 12.31). WHODAS scores above 45 are considered indicative of severe functioning problems (Table 17). Table 16 disaggregates the health condition and the disability percentage, showing the number of cases and percentages, as well as the mean WHODAS score and its standard deviation. Mean WHODAS scores of above 50 are only exceptional in the two lower disability severity ranking groups.

4.4.2 Comparing disability severity ranking groups: no, moderate, severe, and very severe

Further comparison between WHODAS scores and certified disability percentages requires that one sets the cut-off points for disability severity ranking groups on the WHODAS Rasch score scale of 0-100. To compare a "like for like" in the case of Bulgaria this means determining cut-off points for no, moderate, severe, and very severe disability. (As mentioned above, the current Bulgarian disability assessment system identifies disability levels as percentages – with values < 50 percent designating mild/no disability, 50-70 percent moderate disability, 70-90 percent severe disability, and > 90 percent very severe disability).

Table 17: WHODAS score distributions in the World Bank pilots in Latvia, Lithuania, and Greece and suggested cut-off values.

WHODAS-score range	**Latvia** Rasch-based 0-100 score	**Lithuania** Rasch-based 0-100 score	**Greece** Rasch-based 0-100 score	**Proposed cut points for the Rasch-based 0-100 WHODAS-score**
< (Mean – 1SD)	< 38	< 46.6	< 29.5	Cut-off chosen based on 15% of disability from the WRD[15] and looking at the cut-point in norms from the WHODAS manual. This approach delivers a more accurate cut-point based on general population data. **0 to 25 [label: No disability]**
(Mean – 1SD) to Mean	38 to 47.4	46.6 to 55.1	29.5 to 46.6	**26 to 45 [label: Moderate]** Note that this group includes "mild difficulty - 1" answers to WHODAS questions, hence, this category should be "mild to moderate".
Mean to (Mean + 1SD)	47.4 to 56.6	55.1 to 63.6	46.6 to 63.8	**46 to 60 [label: Severe]**
(Mean + 1SD) to 100	> 56.6	> 63.6	> 63.8	**61 to 100 [label: Very severe]**

[15] World Health Organization & World Bank. (2011). World report on disability 2011. World Health Organization. https://apps.who.int/iris/handle/10665/44575

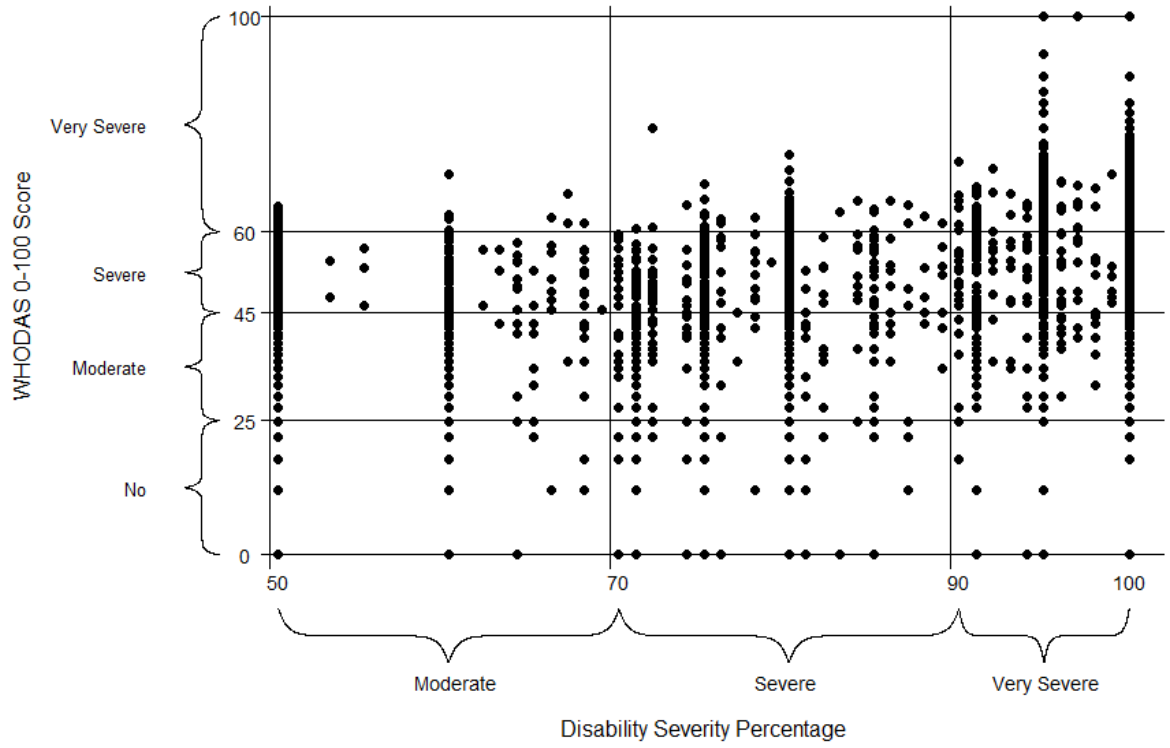
To the best of our knowledge, there are no published cut-offs for the WHODAS score that would be applicable to a population with diverse health conditions to categorize the severity of their disability. Having cut-offs would make it possible to easily detect individuals with significant disabilities and use this information to attribute a disability severity ranking. Some studies report that the 90^{th} even 95^{th} percentile of WHODAS scores would be the best cut-off to determine severe disability in some specific groups, such as post-partum women (Mayrink et al. 2018) or the elderly (Ferrer et al. 2019). In the mixed condition population of this pilot, the 90th and 95th percentile of the WHODAS score was 66 and 70. Minimally clinically important difference scores for the WHODAS have not been established yet (Federici et al. 2017). However, based on several past pilots conducted by the World Bank using WHODAS, in Greece, Latvia, and Lithuania, WHODAS disability cut-offs for the Rasch-based 0-100 score in Bulgaria are suggested as follows:

- 0 – 25 points: No/mild functioning restrictions
- 26 - 45 points: Moderate functioning restrictions
- 46 - 60 points: Severe functioning restrictions
- 61 - 100 points: Very severe functioning restrictions

Table 17 presents the reasoning and score distributions that lead to these WHODAS disability cut-offs for the Rasch-based score on 0-100 scale.

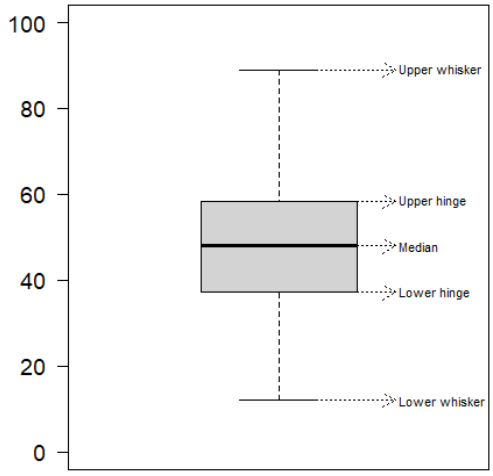
Figure 8 presents a scatter plot of the distribution of the data points for the WHODAS score and the disability percentage. The vertical and horizontal lines represent the cut-offs for the disability severity percentage (vertically) and for the WHODAS score (horizontally). These two perspectives on disability assessment produced measurement values with only a moderate positive correlation of $r = 0.45$.

Figure *8*: Disability percentages and WHODAS score distribution with respective cut-offs



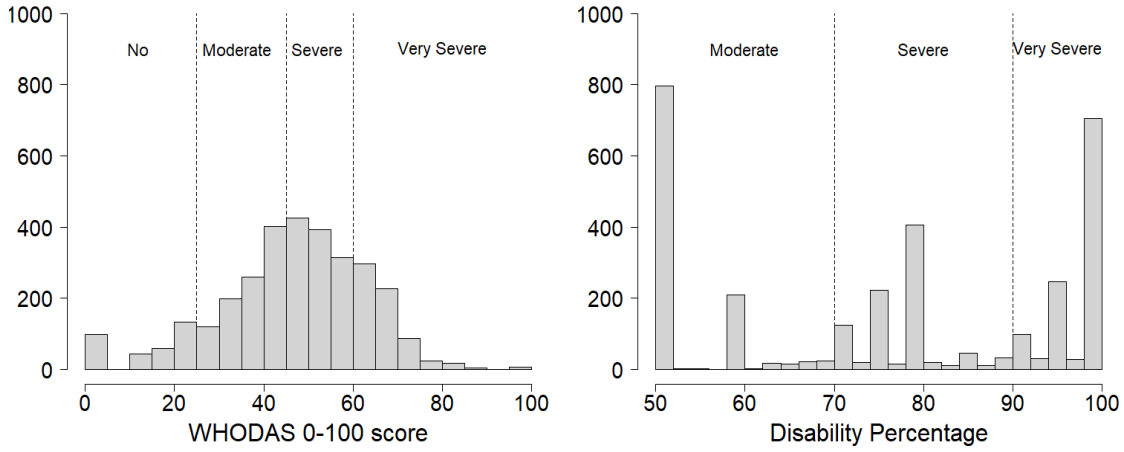
The distribution of the WHODAS scores can also be described in boxplots. The hinges display a confidence interval around the median (50th percentile) and represent the 25th and 75th percentile, or first and third quartile (Q1 and Q3). The whiskers represent the reasonable extremes of the data, i.e., the minimum and maximum values that do not exceed a certain distance from the median (Figure 9).

Figure 9: Boxplot – Terminology



The distribution of the reduction or disability severity rankings and WHODAS scores are radically different: the reduction percentages, first start at 50 percent and polarize on a few locations of the 0-100 percent continuum, i.e., 50 percent (N = 797), 80 percent (N = 404), and 100 percent (N = 699) (Figure 10). By contrast, the score distribution of WHODAS is statistically normal, with a mean score of 46.7 and standard deviation of 16.6. The mean score, standard deviation, and quartiles of the WHODAS score distributions are shown in Table 12.

Figure 10: WHODAS scores and certified disability percentage distribution by respective cut-offs



The results presented above come as no surprise as WHODAS was designed explicitly to assess whole-person disability, while the medical approach to assessing disability used in Bulgaria, does not directly

assess disability, but infers disability based on the underlying health condition or impairment. Sometimes there is a very close correlation between severity of health condition and severity of disability; but sometimes there is no connection. We clearly see this in the case of mental health problems where the impact of the person's environment may greatly increase the impact of the experience of, e.g., depression. This is the basic validity problem with medically based disability assessment. As pointed out above, although the presence of a health condition and associated impairments is a precondition for disability, inferring the level of disability from the presence of a health condition is scientifically problematic. The level of disability that an individual experiences, as the ICF argues, is determined by an interaction between a health condition and associated impairments and environment in which the person lives. WHODAS was designed to directly capture this disability experience while assessment of disability based solely on medical grounds cannot do so validly or reliably.

We have noted above that the assessment method used in Bulgaria is based on medical conditions and associated impairments and pre-determined percentages of disability/ – the so called Barème method. It is significant that concerns about the Barème method are widely held in both the scientific and policy communities. A good and clear example of these concerns in the specific case of disability assessment is a Council of Europe's report on disability assessment in Europe, published in 2002.[16] In this report, the Barème method is characterized as "an arbitrary ordinal scale which attaches progressive percentage values to define disabilities. The disabilities of the claimant are compared to those for which there are scale values, and a percentage is thereby obtained".[17] The Report goes on to observe that "We had no information on the reasons for choosing the levels set out in the Baremas. It seemed that both social and medical factors had been considered. In at least some cases, there seemed to be no mechanism for reviewing and updating Baremas in the light of changes in epidemiology and medical progress affecting the management and prognosis of conditions, let alone social pressures on the benefit system. … There was no clear evidence to us of how clinicians applying such scales to make their decisions. … The most difficult area for impairment-based systems is that of mental health problems."

Focusing on the specific challenge of using the Baremic approach to create summary values, essential for disability assessment, the report states:

"There are a number of inherent problems in deriving a single summary figure for awards based on Baremas, which may be summarized as:

> *The Set Points*: how do you compare a fractured leg with schizophrenia, without giving ranges of values for one or the other which are really little guide to the user?
>
> *The Paired Organs Problem*: what do you do about the one-eyed man who loses it?
>
> *The Whole-Body Problem*: if loss of a finger is 10 percent, and back pain is 20 percent, and depression is 40 percent, what is the total score for an individual with all three conditions?

[16] Council of Europe. 2002. *Assessing Disability in Europe, Similarities and Differences*, Report drawn up by the Working Group on the assessment of person-related criteria for allowances and personal assistance for people with disabilities (Partial Agreement) (P-RR-ECA). Integration of people with disabilities. https://rm.coe.int/16805a2a27/.

[17] Ibid. Scales of compensation for injuries date back at least to mediaeval times in Europe. They related sums for loss of body parts to the 'wergeld' or 'manngeld' (sum payable as a compensation for the killing of a free man) found in Germanic law. The method of transposing such scales to percentages was introduced by the French mathematician François Barrême in late XVIII Century. They have therefore become known as Baremas.

The Threshold Problem: if eligibility for benefits is awarded at a threshold (such as 30 percent for a partial disability pension, and 80 percent for a full one), how do you decide whether someone falls at 29 percent, 30 percent or 31 percent?

The report concludes that "In most cases, it is doctors who apply Baremas, because: it was the custom when these early systems were evolved, they are generally seen as difficult to apply, so a highly qualified examiner is needed, and their application is very reliant upon the results of examinations and tests of which doctors have thorough knowledge".[18]

We agree with this analysis; and the results of this pilot merely confirm these concerns and the specific problems that the Council of Europe Report has focused on when it comes to relying wholly on a medical Baremic approach, as is used in Bulgaria and indeed in many other countries in Europe, without relying on directly obtained functioning information, such as the information collected with the WHODAS questionnaire.

4.4.3 Real life examples

To illustrate the discussion above, we present 6 randomly selected six real life cases from the WHODAS pilot data set where disability percentage and WHODAS scores differ dramatically (also summarized in Table 18).

Case A is a 62-year-old divorced man with a WHODAS-based functioning score of 63, which would indicate very severe functioning restrictions. He has been determined having a moderate disability severity of 66 percent. He reports 11 years of education and lives independently in the community but cannot work for health reasons. His main condition is an unspecified cirrhosis of liver, but he further presents a personality disorder and hypertensive heart disease. He reports also having had difficulties because of his health condition on every day of the last month. He is unable to perform usual activities and often must reduce usual activities or work.

Case B is a 59-year-old married woman with 15 years of education. She is currently working. She suffers an unspecified cirrhosis of liver accompanied by some anemia, a hip arthrosis, and gout. Her reduction has been rated as very severe, i.e., 93 percent. Her WHODAS-score of 36 indicates that she has only moderate functioning problems in a day-to-day life. She also reports marginal difficulties in carrying out her activities and work, having to reduce or cut-back activities only about one day per month, when she is not feeling too well.

Case C is an 84-year-old married man with a WHODAS-based functioning score of 79. He has been determined having a 'severe' disability with a percentage of 72 due to a Type 2 Diabetes and a heart failure. He is retired but still lives independently in the community. His health condition is severely limiting him in his daily life, and he cannot perform his usual activities normally without having to reduce them.

Case D is a 45-year-old educated and working woman. She has never been married. She was diagnosed with diabetes mellitus with multiple complications, including hypertension and chronic pancreas problems. Her disability has been rated as very severe, i.e., 94 percent. Based on the ratings of the WHODAS, she reports only moderate functioning problems in daily life, her score being 27. She also reports only marginal difficulties in carrying out her activities and work, she is never totally unable to carry out her work or activities because of her conditions and only must slow down somewhat from time to time.

[18] Ibid.

Case E is a 44-year-old married woman with 17 years of education. She is unemployed but not for health reasons. She is in the severe disability severity group with a percentage of 85. She has been diagnosed with an organic personality disorder without further comorbidities. Her WHODAS-score of 22 indicates good functioning and her health condition is not limiting him in performing daily life activities.

Case F is a 19-year-old married man with 12 years of education living independently in the community. He is unemployed for health reasons. He is in the moderate disability severity group with a percentage of 50. He has been diagnosed with a mild mental retardation without further comorbidities. His WHODAS-score of 60 is high and indicates severe functioning problems with his health condition limiting him every day of the month, having to reduce activities half of the time.

Table 18: Disability percentages and severity grouping and WHODAS scores and severity grouping - real life cases

Case	Disability % and group	WHODAS score and group
Case A	66% - moderate	63 – very severe
Case B	93% - very severe	36 - moderate
Case C	72% - severe	79 – very severe
Case D	94% - very severe	27 - moderate
Case E	85% - very severe	22 – no difficulty
Case F	50% - moderate	60 – very severe

The cases presented above corroborate the above discussion that the assessment based on medical information may misrepresent the true extent of disability an individual experiences. This is very important, because an accurate assessment of disability is crucial for persons experiencing disability to access disability benefits. For example, cases A, C, and F will have no access to personal assistants (their disability percentage is less than the threshold of 90 percent), although they experience severe difficulties in functioning. In contrast, cases B and D will be eligible for personal assistance, although their disability experience in terms of functioning is moderate. Including functioning into disability assessment in Bulgaria will, thus, not only improve the accuracy of the extent of disability assessment but will also improve the assessment of needs of persons with disabilities.

4.5 To conclude

The empirical evidence presented in the sections above shows that:

- WHODAS – is a freely available and widely used questionnaire built on the activity and participation domains of the WHO's International Classification of Functioning, and Health, that is as close to being the gold standard for the description of disability as we have. It is psychometrically strong, and the data can be analyzed to create a valid and reliable interval-scaled functioning score. This evidence from the Bulgarian pilot corroborates evidence from other international research studies: WHODAS successfully collects functioning information, and as it has been further confirmed by pilot data, it does so with strong psychometric properties of validity and reliability. It performs well in measuring whole-person disability, creates a summary score, and provides an objective and accurate assessment of functioning based on core functioning domains of the ICF. The scores provide interval-scales values ranging normally from 0 to 100 (Figure 9).

- The current system that determines the disability severity ranking groups in Bulgaria has well-known scientific concerns about its capacity to validly assess the true extent of disability an individual experiences and in particular it does not differentiate degrees of disability accurately. The percentage continuum from 0 to 100 is poorly populated and polarizes on a few values (Figure 10).
- In light of these results from the pilot, we conclude that including the assessment of functioning based on WHODAS[19] into the assessment already in place in Bulgaria will significantly improve the accuracy of the assessment, resulting in a more refined assessment that adds information on the lived experience of the disability to the assessment based on the medical diagnosis.

5. Options for including functioning into disability assessment in Bulgaria

5.1 Methodological considerations

The WHODAS pilot in Bulgaria has shown that it performs well in capturing disability experience. The question is how best to include the functioning information captured by WHODAS into the disability status assessment system in Bulgaria.

It should be emphasized that we are not suggesting that medical information, or even assessment of kind and degree of disability based on medical information, should play no role in disability assessment in Bulgaria. This is clearly not the case. The ICF itself makes it clear that without an underlying health condition and associated impairments, disability does not exist, so medical information will also be relevant to disability assessment. While we, and the scientific community, find problematic the direct inference of whole person disability from medical information alone, this information must still be collected and relied on for disability assessment. Information about health states provide a basis for identifying specific physical and mental dimensions of activities and areas of participation that are vulnerable to disability, which can then be directly confirmed by WHODAS data. Medical information provides essential guidance on the medium- and long-term trajectory of disability that the individual will experience, including whether the person faces a progressive decline in health capacity resulting in more and more disability, or the reverse, a progressive improvement. In short, medical information is an essential component of disability assessment.

As medical information is essential, in this section of the Report, we analyze and discuss possible options for combining medical and functioning information in the assessment of disability in Bulgaria.

As we have done in other countries, several methods were tested on the Bulgarian pilot dataset to address this challenge. These methods can be grouped here into two principal strategies (1) ***averaging*** the medical assessment percentage with the WHODAS score to arrive at a final or disability assessment score, and (2) ***flagging*** persons whose WHODAS score, and disability severity group are different from the severity group based on the percentage determined based on medical information.[20]

[19] We recommend WHODAS, because it is free and firmly empirically proven that it represents the construct of disability in terms of ICF and is psychometrically valid and reliable. Countries may choose to develop their own instruments, but such effort requires time and money, and the instrument will have to be psychometrically tested before being deployed.

[20] It is important to add that as WHODAS is used more and more data are collected, this data can be further analyzed using the techniques from this Note to continually update and recalibrate parameters and cut points. Moreover, these data have other potential policy applications, including in identifying disability trends and planning for the future.

(1) **Averaging** – averaging the attributed disability percentage and WHODAS score. This approach is based on the theory that, together, medical, and functioning scores contribute, to different degrees, to a realistic and valid assessment of disability. Below we show the results of four strategies (#2 - #5) that were tested using different weighting combinations.

(2) **Flagging** – identifying persons whose WHODAS severity grouping differs from the medically determined severity grouping and flagging these individuals to request from them additional information or reassessment, or otherwise altering the overall disability percentage to account for the reported level of functioning. Strategies #6 to #8 represent different flagging scenarios, based on the suggested WHODAS severity grouping cut-off values. The flagging approach assumes that medical information may misrepresent the true extent of disability an individual experiences. When an individual has a WHODAS score over or below some cut-off, this suggests that the medical score does not adequately capture the experience of disability and additional information and second level assessment is required.

Averaging and Flagging are, arguably, the most intuitively obvious approaches to merging diverse assessments into a single overall assessment. Each is grounded in the ICF understanding of disability as the outcome of an interaction between the underlying health condition and impairments of a person and the physical, human-built, interpersonal, attitudinal, social, economic, and political environment in which the person lives and acts. They differ, however, in how they weigh the impact of the medical and environmental determinants of disability.

Table 19 gives an overview of the testing strategies that were considered and gives the number of individuals that would be considered having a moderate, severe, or very severe disability after adjusting for the WHODAS-score. Further, the number of individuals that would have their disability severity ranking group changed towards a higher or a lower group are shown. Strategy #1 was included as the current situation in which the basic medically based disability severity percentage is used (and where the WHODAS has 0% effect on the scoring).

Table 19: Overview of WHODAS inclusion strategies

General Approach	Nbr.	Description of scores integration formula	Cut-off	Total Moderate Disability	Total Severe Disability	Total Very Severe Disability	Total Upshift Disability	Total Downshift Disability
Actual approach	#1	Reduced Working Capacity [%Disability*]	• No disability < 50% • Moderate 50-70% • Severe 70-90% • Very Severe > 90%	1096	912	1110	0	0
Averaging:	#2	Weighted mean of %Disability (75%) and WHODAS (25%)	Bivariate cut-offs for %Disability cut-offs and equivalent critical WHODAS Score level	1235	938	945	30	334
	#3	Weighted mean of %Disability (50%) and WHODAS (50%)		1342	998	778	74	652
	#4	Weighted mean of %Disability (25%) and WHODAS (75%)		1346	1080	692	224	892
	#5	Weighted mean of %Disability (0%) and WHODAS (100%)		1315	1133	670	552	1211
Flagging:	#6	WHODAS-Score = Severe disability & %Disability = Moderate		1053	955	1110	437	0
	#7	WHODAS-Score = Ver severe disability & %Disability = Moderate or severe		1090	893	1135	31	0
	#8	WHODAS-score = No to mild disability and %Disability = Severe or very severe		1258	805	1055	0	217

**% Disability stands for the medically attributed disability percentage or work capacity reduction percentage*

The averaging strategies #2 to #5 aggregate medical score and the WHODAS score by giving an increasingly higher weight (25%, 50%, 75%, 100%) to the WHODAS score. Having two cut-offs for the disability percentage, two lines are rotated to modify the division of disability percentage spaces, these two lines being set at the threshold from moderate to severe and the threshold from severe to very severe percentage group. The lines rotate around the WHODAS cut-off values for the

corresponding thresholds (moderate-severe and severe-very severe) at 45 and 60 respectively. The Figures 11 to 15 visualize the averaging procedure and how the disability group division vary with increasing weight of the WHODAS.

The flagging strategies #6 and #8 use the WHODAS-score distribution to determine extreme scores and the cut-offs are either population-based or specific to each disability percentage group. Strategies #6 and #7 flag participants in disability severity ranking groups that have a WHODAS-score indicating a higher level of disability. Figures 16 to 18 visualize the flagging approach and illustrate also how the health condition can impact the score distribution.

5.2 Options

Below, we present options to include functioning into the disability assessment. Each option follows the ICF theory in as much as it combines the medical component of assessment with a functioning component, assessed by WHODAS. Option A is the situation in which WHODAS scores are considered in a purely discretionary manner. Options B, and C are quantitative strategies. Each of these assessment options is described below, with advantages and disadvantages of each. Our framework for evaluating these options – based on the scientific literature – are key scientific principles that determine the credibility of any disability or assessment process: validity (the extent to which the option relies on a true assessment of disability); reliability (the ability of the option to arrive at the same assessment of the same case by different assessors); transparency (the degree to which the assessment process and outcomes can be described and understood by all stakeholders); and standardization (the extent to which the process resists distortion or alteration over time and across locations).

Option A: Discretionary combination of medical and functioning components

This is the option in which an individual or committee reviews medical scores and the WHODAS scores and makes a judgment about the extent of disability as the individual or committee sees fit. This is a purely discretionary option, and it is surprisingly common in practice. As an option for disability, it has the (minimal) advantage of simplicity and administrative convenience. On the disadvantage side, however, this approach is subject to manipulation, or whim, lacks validity and reliability, and is utterly non-transparent. The option is given here in part as a contrast to the remaining options B and C, but also, in fairness, because some countries continue to rely on this option for disability assessment. We do not recommend this option.[21]

Options B and C: quantitative approach

The two remaining options are quantitatively driven, which makes them very different from Option A. In different ways and for different reasons, the two options satisfy not only the basic psychometric properties of validity and reliability but each, to different degrees, strive to achieve transparency and standardization. Two preliminary technical points should be kept in mind:

1) The baseline - namely Strategy #1 –uses only the ICD health condition information to determine the reduction percentage, with three cut-offs 50.0 percent, 70.0 percent, and 90.0 percent, for determining eligibility for benefits. Strategy #1 is the approach that was applied. The pilot sample

[21] Interactions with officers involved in disability assessment in different countries suggest that medical professionals involved in the assessment disability are confident they "know best" and can consider functioning and the experience of disability as part of the medical description of the applicant's situation. One often hears medical assessors claim that they take functioning fully into account when examining medical records. One implicit result from the pilot is that this assumption is not grounded in evidence.

included only individuals with at least moderate disability: 1,096 respondents were attributed a moderate, 912 a severe and 1,110 a very severe disability.

2) As noted above, we posit the Rasch-based WHODAS scores 25, 45, and 60 as the cut-off for moderate, severe, and very severe disability. While moderate disability entitles individuals to some benefits in Bulgaria, the menu and amount of support interventions increase with the degree of disability with very severe (>90.0 percent) with most benefits, including personal assistance. Scientifically speaking, it is essential to create some cut-off values since there is no 'gold standard' for when disability is significant. Ultimately, the cut-off is a socio-political decision that should be transparent and evidence-based in the sense that it represents a plausible threshold based on an analysis of disability prevalence in a population.

Option B: Using an averaging algorithm

In the Bulgaria pilot WHODAS data set, there is a relatively high percentage of persons indicating no functioning problems at all (10.71%), among which some individuals were in the very severe disability severity ranking group (9.28%). Averaging the disability percentages with the WHODAS score would somewhat adjust the number of persons in each of the disability severity ranking groups by accounting to some degree for the observed disability level assessed by the WHODAS. To get a full sense of the range of possible approaches under Option B, four weighting schemes were tested when creating the 8 strategies:

- 75.0 percent disability percentage & 25.0 percent WHODAS score,
- 50.0 percent disability percentage & 50.0 percent WHODAS score,
- 25.0 percent disability percentage & 75.0 percent WHODAS score,
- 0.0 percent disability percentage & 100.0 percent WHODAS score.

There are, of course, many approaches to weighting that might be adopted (and any other arrangement can be constructed, and its consequences determined using the same methods as used for these four), but these four are perhaps the most intuitive.

In this pilot, given the presence of more than one cut-off value for the medically determined disability percentage, the averaging affects the two cut-off lines that separate moderate from severe and severe from very severe disability levels. Further, the WHODAS cut-off values that discriminate these same degrees of disability are used as center point for the rotation of the cut-off lines:

- Moderate to Severe point: disability cutoff at 70 percent and WHODAS score of 45 points
- Severe to very severe point: disability cutoff at 90 percent and WHODAS score of 60 points.

Advantages of Option B:

- An assessment of the level of functioning plays a significant role in the determination of eligibility for disability benefits so that the eligibility for benefits is not solely based on purely medical criteria.
- The averaging approach minimalizes the impact of the inherent psychometric problems with the disability percentage based on the Baremic medical assessment used.
- The assessment of the level of functioning is empirically and statistically verified.
- This option yields high levels of validity and reliability.
- Merging the results of two assessments scaled by means of 'weighted averaging' is fully objective, transparent, and non-discretionary.
- The method is not sample-dependent.

Disadvantages of Option B:

- There are, potentially, an infinite number of combinations of weighting schemes (i.e., 'strategies'), each of which affects the set of eligible applicants differently and has different budgetary and political consequences. This is an unavoidable fact about the nature of disability as a continuum and the fact that there is not yet scientifically verified or objective cut-off of severity for eligibility.
- Any strategy selected will be objectionable to individuals who, under that strategy, will not be certified as disabled and thus not eligible for any benefits. This signals the need for clear and transparent information dissemination and a solid grievance redress system that may include using tools for clinical testing and determination of functioning, such as ClinFIT20[22] or other tools used or recommended by rehabilitation specialists. It should also be noted that any new method adopted should apply to new applicants only. To smooth the transition, disability recertification may be staged over several years.

Option C: Using the flagging algorithm

The flagging strategy takes advantage of the disability cut-offs shown in Table 17. Strategies #6 and #7 flag participants in the moderate disability group that have a WHODAS-score indicating more than moderate disability or participants in the moderate and severe disability group having WHODAS-scores indicating more than severe disability. Strategy #8 flags, on the other hand, participants that have a severe or very severe disability severity group, but no or very mild functioning problems based on the WHODAS. The disability percentage of these persons, that indicate no functioning limitation in daily life, would also need to be reconsidered.

Table 19 shows that setting the cut-off for the WHODAS score at 45, i.e., having severe to very severe functioning problems, and flagging all those individuals with a disability severity percentage assessed as moderate (i.e., strategy #6) would result in many individuals (N = 437), whose disability severity in terms of functioning would have to be increased to the severe or very severe. Only flagging those with very severe functioning problems (WHODAS score of at least 60) who are in the moderate and severe disability ranking groups results in only 31 persons whose disability severity may be reassessed and augmented. This reiterates the point raised above that the current disability assessment method does not discriminate well between degrees of disability. On the other hand, a relatively large number of individuals (N = 217) in the severe and very severe disability ranking groups presented no disability in terms of the WHODAS scores. From those, almost half had neoplasms as main diagnosis.

Advantages of Option C:

- Scientifically robust and based on actual data.
- Shows that the purely medical approach to disability assessment may not accurately assess disability in many cases – in which, as reported in the WHODAS score, a person is experiencing more/fewer functioning problems in their lives than what the health condition/impairment is thought to imply.
- High levels of validity and reliability.

Disadvantages of Option C:

- The WHODAS cut-offs were recommendations based on past pilots and some evidence from the scientific literature. Sensitivity analyses are not available to this point. More precise cut-values specific to Bulgaria may be introduced at later time points when more information on functioning is collected (assuming that WHODAS will be introduced in Bulgaria).

[22] ClinFIT20 is the official disability assessment tool of the International Society of Physical and Rehabilitation Medicine (ISPRM). See ClinFIT: ISPRM's Universal Functioning Information Tool based on the WHO's ICF, available at: http://www.jisprm.org, IP: 62.98.194.95.

- The flagging approach is vulnerable to political manipulation as the criteria for determining which individuals to 'flag' is discretionary. To resolve this, a robust technical and procedural guidance will need to be in place to direct the second level assessment, to ensure methodological soundness and transparency.

Even with the caveat concerning the cut-off points for disability severity, the flagging method may be introduced through specifically designed (two step) administrative procedure. See 5.4 below.

5.3 Real life examples and illustration of the functioning inclusion options and their impact

The options presented above may seem abstract. To make them concrete, we will illustrate them on the 6 real life cases presented above. This should show how the strategies would change the understanding of the level of disability and highlight the advantage to including functioning into the current disability assessment in Bulgaria.

How would these six individuals be assessed with the combined disability percentage and the (Rasch-adjusted) WHODAS score in the eleven strategies? Table 20 presents the expected level of disability given each of adjustment strategies (yellow = moderate; orange = severe; red = very severe).

Table 20: Disability severity ranking and WHODAS scores and their integration strategies – Examples of individual cases

			Current method severity	**Averaging**				**Flagging**		
	WHODAS score	**Disability percent**	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**
A	63	66						**Additional information and second step assessment**		
B	36	93								
C	79	72								
D	27	94								
E	22	85								
F	60	50								

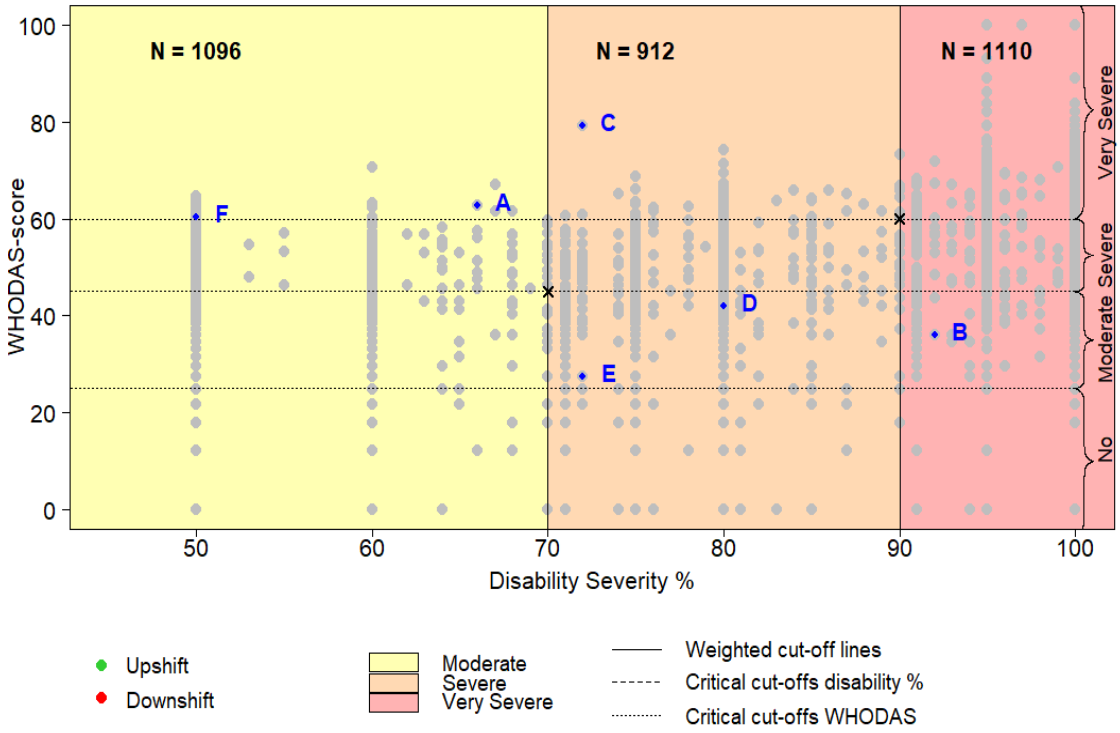
What follows will illustrate graphically how the averaging options function with five relative weightings of the disability percentage and the WHODAS score – weighting the disability percentage at 100 percent and WHODAS at 0 percent (i.e., the actual approach or Strategy #1); weighting disability percentage at 75.0 percent and WHODAS score at 25.0 percent (Strategy #2); and so on. The averaging approach can be easily depicted by the mean of a cartesian coordinate system with the disability percentage on the x-axis and the WHODAS score on the y-axis. The two weighted cut-off lines separates between moderate and severe as well as severe and very severe levels of disability. Like a clock hand, the separation line moves with increasing weight of the WHODAS with individuals who are either shifted upward in regions of higher disability or downwards, towards less functioning problems or disability. The coordinate system approach could be implemented in practice to actually 'locate' specific individuals on the graph, based on their disability percentage and WHODAS scores. This makes it possible to see if the scores of an individual are congruent or if there is a discrepancy between the ratings that would require a reiteration of the case and finally re-attribution of the disability severity. For concreteness as well, the six described individuals, A, B, C, D, E, and F are marked in each graph.

Current assessment system

STRATEGY #1 - Disability Severity 100% and WHODAS 0%

Strategy #1 (Figure 11) only considers the disability severity ranking percentage, the cartesian field is divided vertically at a cut-off of 70 percent and 90 percent, with individuals having a moderate disability in the yellow field, individuals with severe disability in the orange field, and individuals with a very severe disability in the red field. This coloring scheme will be used in all Figures illustrating the averaging approach.

Figure 11: Weighted Average Disability Percentage 100% and WHODAS-score 0%



Without any adjustment to the actual approach, the cases A and F would have a moderate disability, Cases C, D, and E a severe disability, and Case B a very severe disability.

Averaging

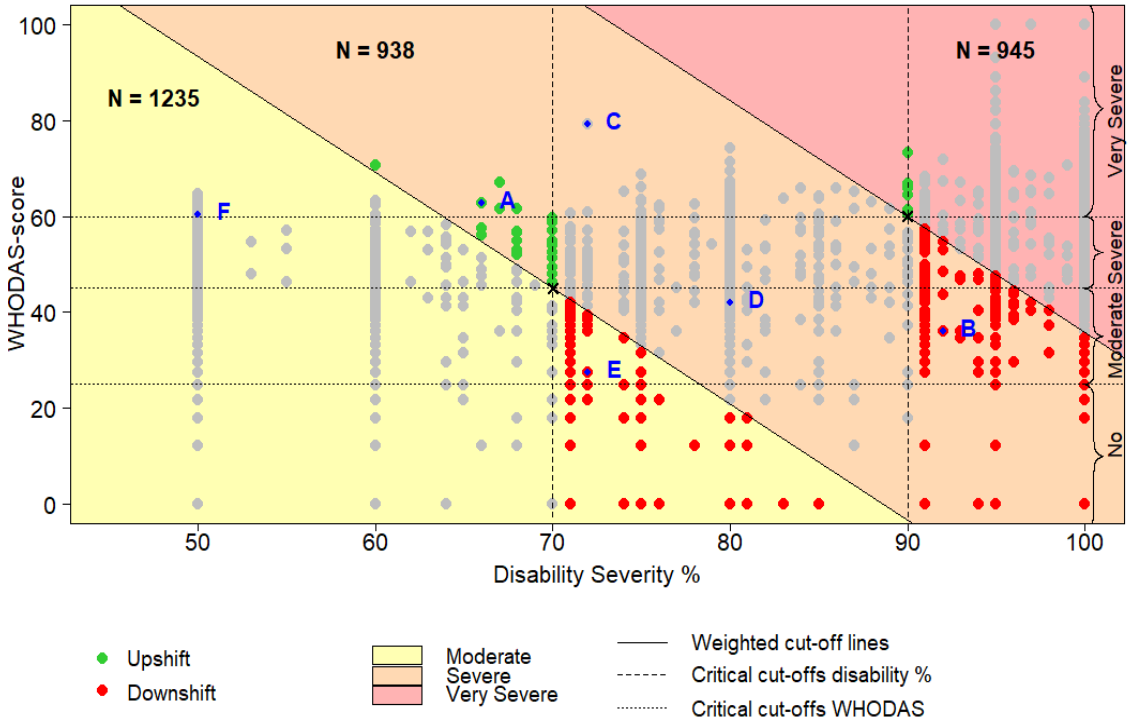
In the averaging strategies #2 to #5, the cut-off lines are gradually rotated around cut-off points for the WHODAS scores that separate moderate from severe functioning problems (WHODAS-score = 45) and severe from very severe functioning problems (WHODAS-score = 60).

STRATEGY #2 (Figure 12)

Disability Severity 75% and WHODAS 25%, with WHODAS cut-off at median score.

In strategy #2, WHODAS contributes 25.0 percent to the disability assessment, this would change the disability level of Cases A and E. Case A has moderate disability and very severe functioning problem. By including functioning into the disability assessment, his case would need to be reconsidered for a higher disability level. Case E has very severe disability and low functioning score. Adding functioning information to the equation would reduce the degree of her disability from severe to moderate. Applying this strategy would upshift N = 30 individuals towards more disability and downshift N = 334 individuals towards less disability.

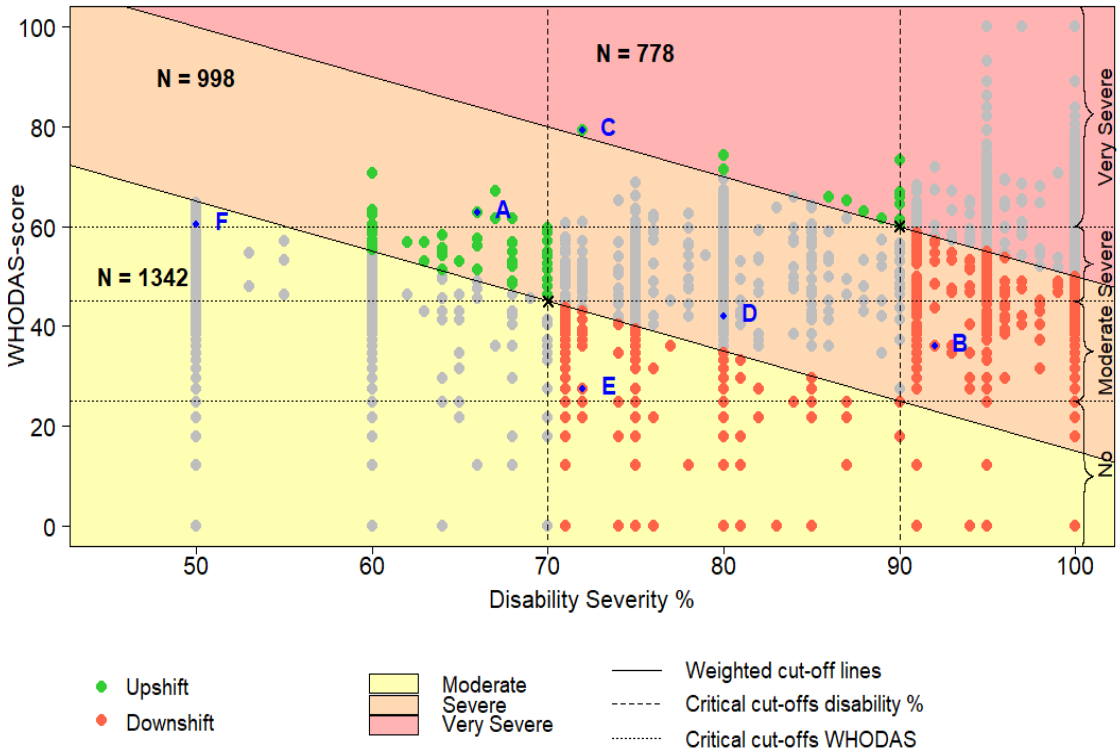
Figure 12: Weighted Average Disability Percentage 75% and WHODAS-score 25%



STRATEGY #3 (Figure 13)

Disability Severity 50% and WHODAS 50% with WHODAS cut-off at median score

Figure 13: Weighted Average Disability Percentage 50% and WHODAS-score 50%

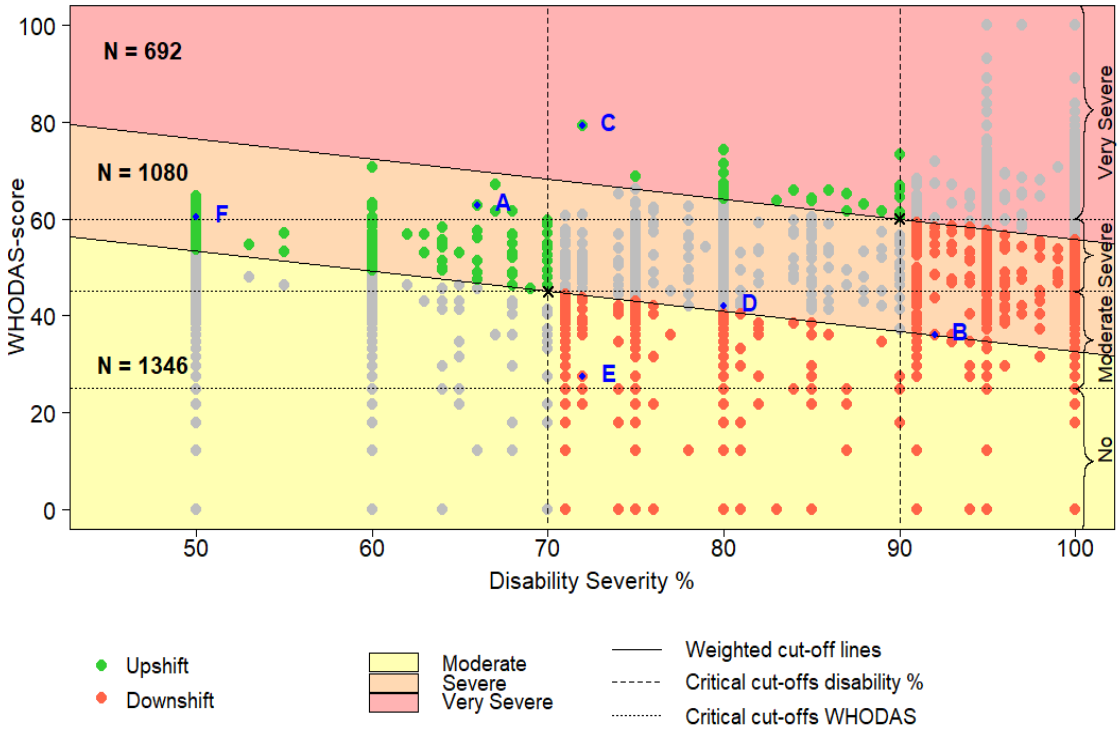


In strategy #3, WHODAS contributes 50.0 percent to the disability assessment, this would change the disability level of Case C. C is a senior man of 84 years with diabetes and a heart failure. He is retired and still living in the community, his health condition is limiting him. Accounting for functioning, his case would change from severely to very severely disabled. His WHODAS-based functioning score of 79 is very high and indicates severe limitations in daily functioning. Applying this strategy would result in shifting N = 74 individuals towards more disability and N = 652 individuals towards lesser disability.

STRATEGY #4 (Figure 14)

Disability Severity 25% and WHODAS 75%, with WHODAS cut-off at median score

Figure 14: Weighted Average Disability Percentage (25%) and WHODAS-score (75%)

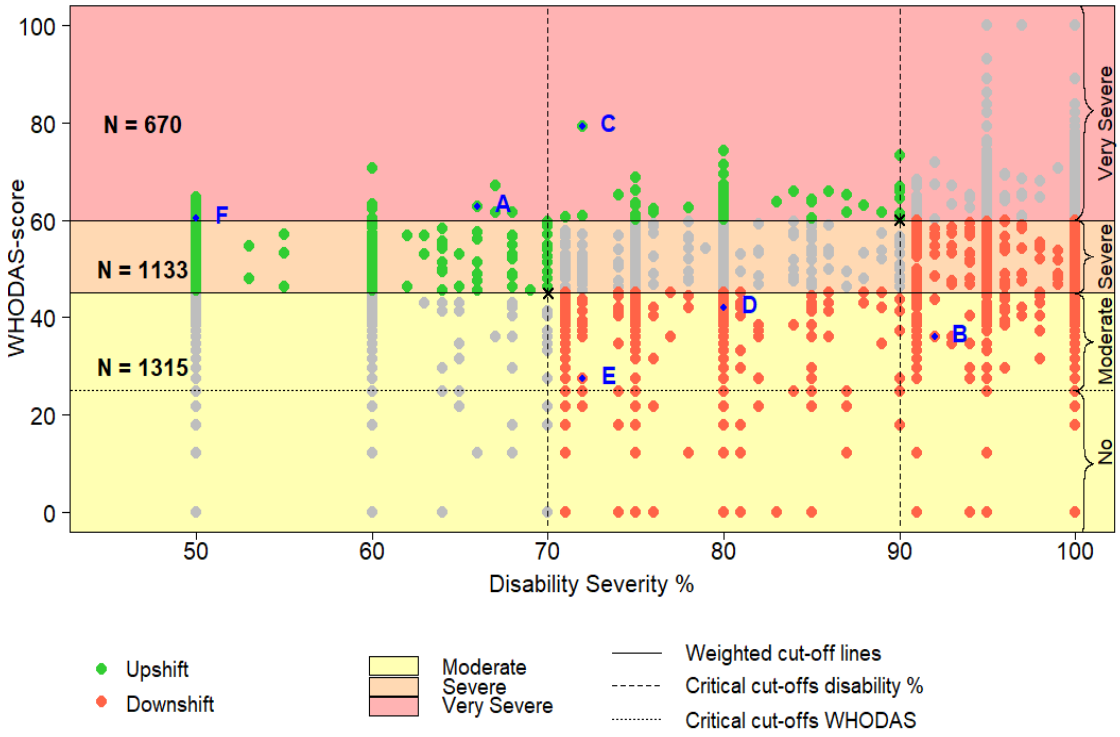


In strategy #4, the functioning assessment would receive more weight, with WHODAS contributing 75.0 percent to the disability assessment. In this strategy, Case F would be upshifted towards more disability. F is a 19-year-old married man diagnosed with mild mental retardation. His WHODAS score is high and indicates severe functioning problems limiting him on a daily basis. He was attributed a moderate disability percentage but based on his high WHODAS-score his case would need some clarification. Applying this strategy would result in shifting N = 224 individuals towards more disability and N = 892 individuals towards lesser disability.

STRATEGY #5 (Figure 15)

Disability Severity 0% and WHODAS 100% with WHODAS cut-off at median score

Figure 15: Severe Disability based on WHODAS-scores



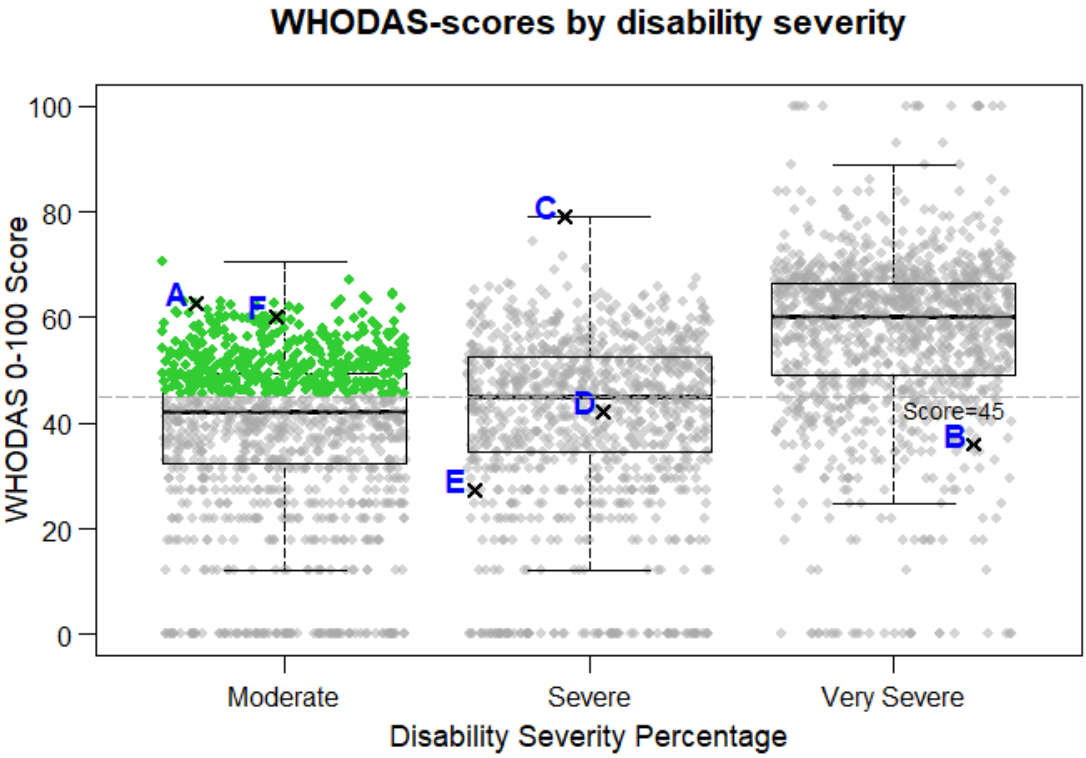
Strategy #5 is the most extreme strategy at it would only account for the functioning information derived based on the WHODAS. Cases A, C, and F would need to be reconsidered, given their high WHODAS-score that indicates very severe functioning problems. On the other hand, Cases B, D, and E would indicate a moderate disability and not a severe or very severe disability, as indicated by the reduction percentage they were attributed.

Flagging

Flagging would only highlight individuals where the WHODAS-score and disability severity grouping differs from medically determined disability grouping.

STRATEGY #6 (Figure 16)

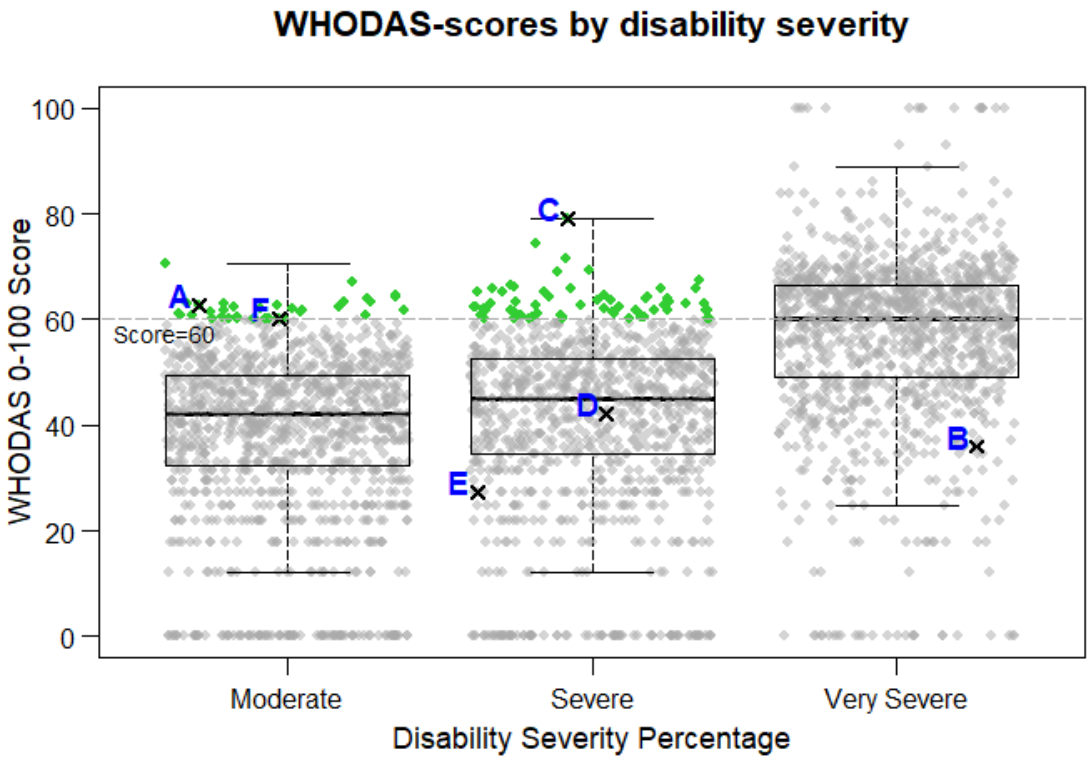
Figure 16: Severe Disability based on WHODAS-scores



Strategy #6 flags and may benefit individuals that have a WHODAS score above 45, i.e., individuals expected to have severe to very severe functioning problems, but who have been linked to the moderate disability ranking group. These persons – the green dots in the Figure 17 - should be reconsidered, as their disability percentage may not entirely capture the functioning problems that they experience in daily life. With this strategy, Cases A and F may shift upwards to a severe disability level. With this strategy, which requires a WHODAS-score of 45 and a moderate disability percentage, N = 437 individuals would be flagged as having more than moderate functioning impairments.

STRATEGY #7 (Figure 17)

Figure 17: Very Severe Disability based on WHODAS-scores



Strategy #7 flags and may benefit individuals that have a WHODAS score above 60, i.e., individuals with very severe functioning problems but who have been attributed only a moderate or severe disability percentage. Here also, the individuals that could be reassessed for an upshift are colored in green. With this strategy, which requires a higher WHODAS-score, N = 31 individuals would need to be reconsidered and discussed in light of their reported functioning limitations, including Cases A, F, and C.

STRATEGY #8 (Figure 18)

Figure 18: No Disability based on WHODAS-scores

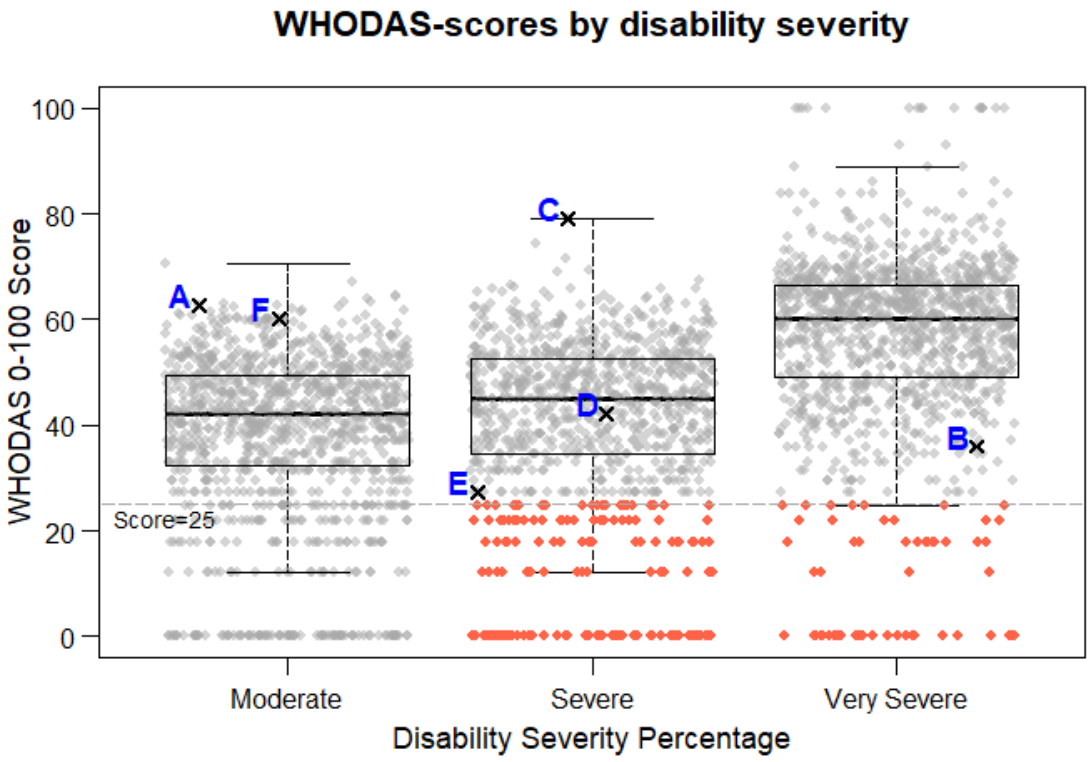


Figure 18 indicates that flagging could also be used, not only to flag persons with very high levels of disability that could be missed by the system in place, but also to flag individuals with an unexpected low disability level compared to their severe or very severe disability severity ranking. WHODAS scores below 25 indicate a low level of disability, i.e., that is not expected to impair a person in daily functioning. A total of N = 217 individuals whose score indicates no functioning problems, being in the severe or very severe disability severity ranking groups would be flagged and may be reconsidered.

5.4 Implementation considerations

This Note has provided evidence that the current disability assessment system in Bulgaria would benefit significantly from the inclusion of functioning into the assessment method: (i) the assessment of disability would be more precise and accurate, reflecting the real life experience of disability of applicants; (ii) the assessment will be in line with modern understanding of disability; and (iii) the assessment will be harmonized with the individual needs assessment providing valuable input into it. A status assessment that includes functioning will provide a better profile of disability that the person experiences to identify needs that, once addressed, will improve the experience of disability by optimize the person's functioning.

Our approach to disability assessment is to combine medical and functioning information and we have provided above several methodological options for doing it. An important further question is whether Bulgaria has administrative capacity to implement the change smoothly and without significant cost. The answer is yes. First, Bulgaria has an advanced information system that could easily accommodate the collection and use of the information on functioning (see Chapter 7 in the above-mentioned World Bank report on Disability System and Policy in Bulgaria). Second, Bulgaria has a cadre of experienced

social workers in the Social Assistance Agency that could be engaged in the WHODAS administration. While administrative process will have to be designed and details worked out, it could possibly flow in the following way: a person applying for/referred to the assessment of disability would have two meetings scheduled: one with the social worker to administer WHODAS and subsequently one with the TMEC. The WHODAS information would be sent to NMEC electronically where the form would be checked, and the raw score transformed into the Rasch based score. TMEC will proceed with the assessment as per the current criteria.

How the two scores will then be combined depends on the choice made by the Government. If the averaging method is chosen, say with 50 percent TMEC and 50 percent WHODAS Rasch score, the two scores will automatically be combined by NMEC, and the final score sent to the TMEC to issue the certificate. The certificate can also be issued by NMEC. If the flagging method is chosen, then in cases where the TMEC determined percentage of disability and the WHODAS score fall in the same disability grouping (no, moderate, severe, and very severe), the NMEC will instruct the TMEC to issue the certificate with the proposed disability severity grouping. If they do not coincide, then a secondary assessment is undertaken either by a different TMEC or a NMEC. Whichever the ultimate choice might be, the result is that the information on functioning will be systematically included in disability assessment using a standardized approach, and the administrative process itself will become more rigorous, standardized, and objective.

6. Recommendations

Based on the above, we strongly recommend that the ***Bulgarian Government includes functioning into disability assessment using WHODAS to collect relevant information***. While the choice is political, either averaging or the flagging approach can comfortably be implemented based on the existing information systems and human resources (a cadre of social workers).

Including functioning into disability assessment will:

- make the assessment of disability more precise, accurate and reliable, reflecting the real-life experience of disability of applicants,
- bring the assessment closer to modern understanding of disability as formulated by ICF and mandated by CRPD, and
- align it with the individual needs assessment by providing valuable information input into it. A status assessment that includes functioning will provide a better profile of disability that the person experiences to identify needs that, once addressed, will improve the experience of disability by optimize the person's functioning.

We also recommend that a separate assessment tool is developed for children because the tools used for adults are not suitable for children. (WHO does not recommend that WHODAS is used for children and is currently working on a WHODAS instrument for children).

References

Bond, Trevor G., and Christine M. Fox. 2001. Applying the Rasch Model: Fundamental *Measurement in the Human Sciences*. Mahwah, NJ: L. Erlbaum.

Boone, W. J. 2016. "Rasch Analysis for Instrument Development: Why, When, and How?" *CBE Life Sci Educ* 15 (4). https://doi.org/10.1187/cbe.16-04-0148.

Christensen, Karl Bang, Guido Makransky, and Mike Horton. 2017. "Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations." *Applied Psychological Measurement* 41 (3): 178–94. https://doi.org/doi:10.1177/0146621616677520.

Council of Europe. 2002. *Assessing Disability in Europe, Similarities and Differences*, Report drawn up by the Working Group on the assessment of person-related criteria for allowances and personal assistance for people with disabilities (Partial Agreement) (P-RR-ECA). Integration of people with disabilities. https://rm.coe.int/16805a2a27/.

Federici, Stefano, Marco Bracalenti, Fabio Meloni, and Juan V. Luciano. 2017. "World Health Organization Disability Assessment Schedule 2.0: An International Systematic Review." *Disability and Rehabilitation* 39 (23): 2347–80. https://doi.org/10.1080/09638288.2016.1223177.

Fellinghauer, C. S., B. Prodinger, and A. Tennant. 2018. "The Impact of Missing Values and Single Imputation Upon Rasch Analysis Outcomes: A Simulation Study." *J Appl Meas* 19 (1): 1–25.

Ferrer, Michele Lacerda Pereira, Monica Rodrigues Perracini, Flávio Rebustini, and Cassia Maria Buchalla. 2019. "WHODAS 2.0-BO: Normative Data for the Assessment of Disability in Older Adults." *Revista de Saude Publica* 53: 19. https://doi.org/10.11606/S1518-8787.2019053000586.

Holland, P. W., and H. Wainger. 1993. *Differential Item Functioning*. Edited by N. J. Hillsdale. Erlbaum.

Mair, Patrick, Reinhold Hatzinger, and Marco Johannes Maier. 2019. *eRm: Extended Rasch Modeling*.

Marais, Ida. 2013. "Local Dependence." In *Rasch Models in Health*, 111–30. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118574454.ch7.

Masters, Geoff N. 1982. "A Rasch Model for Partial Credit Scoring." *Psychometrika* 47 (June): 149–74.

Mayrink, Jussara, Renato T. Souza, Carla Silveira, José P. Guida, Maria L. Costa, Mary A. Parpinelli, Rodolfo C. Pacagnella, et al. 2018. "Reference Ranges of the WHO Disability Assessment Schedule (WHODAS 2.0) Score and Diagnostic Validity of Its 12-Item Version in Identifying Altered Functioning in Healthy Postpartum Women." *International Journal of Gynecology & Obstetrics* 141 (S1): 48–54. https://doi.org/https://doi.org/10.1002/ijgo.12466.

Nunnally, Jum C., and Ira H. Bernstein. 1994. *Psychometric Theory*. 3rd ed. New York ; London: McGraw-Hill.

Rasch, G. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: [s.n.].

Smith, E. V. 2002. "Detecting and Evaluating the Impact of Multidimensionality Using Item Fit Statistics and Principal Component Analysis of Residuals." *J Appl Meas* 3 (2): 205–31.

Smith, R. M., and C. Y. Miao. 1994. "Assessing Unidimensionality for Rasch Measurement." In *Objective Measurement: Theory into Practice. Volume 2*. Greenwich: Ablex: M. Wilson.

Smith, R. M., R. E. Schumacker, and M. J. Bush. 1998. "Using Item Mean Squares to Evaluate Fit to the Rasch Model." *J Outcome Meas* 2 (1): 66–78.

Stekhoven, D. J., and P. Buhlmann. 2012. "MissForest–Non-Parametric Missing Value Imputation for Mixed-Type Data." *Bioinformatics* 28 (1): 112–18. https://doi.org/10.1093/bioinformatics/btr597.

Posarac, A at al. 2022. *Bulgaria: Disability System and Policy, A Comprehensive Review. © World Bank*

Team, R Core. 2016. "R: A Language and Environment for Statistical Computing."

Tennant, A., and P. G. Conaghan. 2007. "The Rasch Measurement Model in Rheumatology: What Is It and Why Use It? When Should It Be Applied, and What Should One Look for in a Rasch Paper?" *Arthritis Rheum* 57 (8): 1358–62. https://doi.org/10.1002/art.23108.

Ustun, T. B. 2009. *Measuring Health and Disability: Manual for WHO Disability Assessment Schedule (Whodas 2.0)*. Geneva: World Health Organization.

Yen, Wendy M. 1984. "Effects of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model." *Applied Psychological Measurement* 8 (2): 125–45. https://doi.org/10.1177/014662168400800201.

World Health Organization. 2001. International Classification of Functioning, Disability and Health (ICF).

https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health.

World Health Organization. 2010. Measuring Health and Disability. Manual for WHO Disability Assessment Schedule WHODAS 2.0. 2010, Geneva, Switzerland: World Health Organization.

Annex 1 - Bulgaria: Strengthening disability system: WHODAS Pilot Protocol

Background

The Pilot study described below has the overall aim of providing a robust empirical evidence basis to support a recommendation as to the suitable option for an effective and efficient comprehensive individual needs assessment method and process in Bulgaria, grounded in the notion of functioning as described in the World Health Organization's International Classification of Functioning, Disability and Health (WHO ICF). Administratively, a needs assessment depends upon a prior, valid assessment of the level or extent of disability that an individual applicant experiences.

Currently in Bulgaria the information on disability that the SAA receives from the medical expertise commissions (TMECs and NMEC) does not validly assess the overall disability a person experiences as a percentage, because this assessment, based entirely on the Medical Expertise, as was demonstrated in the World Bank report on the Bulgaria's disability system and policies, is not based on functioning information. Functioning information, for the purpose of disability assessment, can only be validly collected by means of a scientifically sound and internationally validated instrument. This information will establish the overall degree of functioning problem (i.e., disability) of a person, which is the first step in allowing for the objective identification of the needs of people with disabilities for the provision of financial and other adequate necessary support. Information on functioning cannot presently be obtained from the current model as it is based on medical expertise and health status of the person.

Because of the current situation in Bulgaria, therefore, to achieve the desired results with respect to comprehensive individual needs assessment, it will also be necessary to pilot test a psychometrically sound functioning tool to align disability assessment with needs assessment. We therefore propose to pilot the WHO instrument WHODAS, in its 36-item, clinically (i.e., by a trained professional) administered version as well. WHODAS is a highly reliable generic instrument developed by WHO to provide a standardized method for measuring disability across cultures and is fully aligned with the International Classification of Functioning, Disability and Health (ICF). As WHODAS was designed to have a range of applications – research, clinical, and administrative – the global experience with WHODAS is extremely wide-ranging. In a 2016 systematic review, nearly 800 references to the application of WHODAS around the world are cited, including in disability contexts.[23] To assess disability according to the model of functioning found in the ICF, in a valid and reliable manner that takes into consideration both problems in functioning, associated with health conditions, and areas of strength in functioning – WHODAS is the only internationally recognized application of ICF to disability assessment.

WHODAS, in either its 12 or 36 question version, or modifications to reflect local needs, has been applied across the world, and in many countries in Europe. Some countries, for example, France, Germany, Switzerland, Belgium, have developed, over the course of several years of development and testing, their own functioning assessment questionnaires. Although doing so is always an option, it is an extremely time- and resource-consuming endeavor which, given the intrinsic scientific soundness of WHODAS, may not be a sensible option for many countries. We recommend WHODAS both because of its scientific power and reliability because it has been used successfully in millions of cases, it is straightforward to train interviewers in the administration of the questionnaire, and these interviewers do not require medical knowledge. WHODAS is therefore the practical, and scientifically sound, choice for disability assessment.

[23] Federici S, et al 2016 World Health Organization disability assessment schedule 2.0: An international systematic review, Disability and Rehabilitation, DOI: 10.1080/09638288.2016.1223177.

The pilot protocol, therefore, describes two, concurrent, piloting exercises: (i) a pilot of WHODAS 36 as an additional component of the current disability assessment process that relies exclusively on the results of the Medical Expertise, and ii) the design and piloting of comprehensive individual needs assessment tool and procedure. These are complementary exercises insofar as the data from the WHODAS pilot will not only provide a true baseline of overall disability level for the pilot population, but it will also provide the basis for analyzing the proposed comprehensive individual needs assessment tool. In both cases the piloting will provide both analytical data on the validity and reliability of both WHODAS and the proposed needs assessment tool, but also information about the administrative feasibility of amending current procedures to accommodate these two assessment processes.

1. PILOTING WHODAS (36-QUESTION VERSION, CLINICALLY ADMINISTERED)

In line with international standards and the overall objective of this project to propose a 'modern approach of evaluating disabilities following the ICF framework', the 36-item, clinically administered (face to face) version of WHODAS will be piloted. Piloting WHODAS has the advantage that data collected from it can be statistically analyzed, using modern techniques such as Rasch Analyses, to produce a linear scale or metric for directly measuring severity of disability across and independently from health conditions. WHODAS results can be statistically compared with the current disability assessment process in Bulgaria that is based on the findings of the medical expertise. As such, WHODAS will provide empirical evidence for the inclusion of functioning into disability assessment, should the Government of Bulgaria use the results of the WHODAS pilot, which will establish a functioning baseline (which currently is absent in Bulgaria) on the basis of which it will then be possible to pursue changes to its current methodology for performing individual needs assessment of support of the people with disabilities (See the Analysis Plan below.)

It must be noted that WHODAS is not itself a needs assessment instrument and was not designed to be. WHODAS was designed to produce a single score that represents the overall, whole-person, level of disability that an individual experiences. By contrast, a needs assessment tool identifies specific disability-related limitations in specific domains (mobility, vision, hearing, cognition) for the purpose of identifies needs for services and supports. However, the validity and reliability of a comprehensive needs assessment tool, of the sort that is requested by the Bulgarian government, requires a prior determination of the level of a person's overall disability, that is, the outcome of a scientifically sound disability assessment. This assessment would also identify residual functioning in domains not impacted by the underlying health condition. This information can be used to provide resources to increase and improve the individual's ability to lead an active life based on the existing functioning levels.

The objectives of the WHODAS pilot

The pilot has two main objectives**:** First, the pilot will demonstrate whether the administration of WHODAS-36 is practically feasible – e.g., whether it is possible to administer a 35-minute questionnaire in a face-to-face format – providing evidence for recommendations on improving administrative processes (with the potential aim of informing an automatic management of information). The risks arising from regulations and the current situation related to anti-pandemic measures in the territory of the Republic of Bulgaria in connection with the prevention of the spread of coronavirus should also be considered, as well as the professional level of the administrative capacity in SAA (with regard to piloting international standard instruments) and workload management.

Secondly, and more importantly, the pilot will collect functioning data by means of a valid, reliable, and psychometrically robust tool (WHODAS), and such data will then be systematically analyzed to

show how: (i) the information can inform the needs assessment process, and (ii) to model potential algorithms for harmonizing medical diagnosis and functioning for an improved assessment of disability.

The pilot, in other words, tests both processes and an instrument. As a process, the pilot will identify obstacles and roadblocks in the procedures in place in the disability assessment and the needs assessment processes. This information can be used to improve administrative procedures for conducting disability assessment and for improving data collection and storage in information services. The pilot will also generate valuable data that will form the basis for a complex statistical analysis that will both confirm the basic psychometric properties of WHODAS-36 (in particular validity and reliability) and to derive recommendations concerning how information on functioning can inform the needs assessment process and how it can be used to augment or refine the current determination of disability status, so that Bulgaria can move towards a disability assessment system based on functioning, should it decide to do so.

Tasks

The following tasks have to be accomplished by the Pilot:

1. Evaluate the practical feasibility of the WHODAS 36-item, clinically (for instance face-to-face) administered version.
2. Analyze the psychometric properties (validity and reliability) of WHODAS using data from pilot.
3. Compare results of the WHODAS with results of current disability assessment performed by the TEMC.
4. Compare information on functioning collected through the current needs assessment instrument with the results from WHODAS.
5. Draw conclusions and make recommendations for the improvement of the existing system, based on the results of the Pilot and in line with good international practice.

Pre-Pilot Preliminary Activities

I. Establishment of Pilot Study Working Group and Pilot resource personnel

MLSP is responsible for all aspects of pilot implementation and monitoring day to day activities. MLSP should also establish a small technical working group of national experts, SAA and others to plan, manage and monitor all implementation details of the pilot. World Bank will provide advice and consultation to the Technical Working Group, and will support the monitoring of the pilot implementation, with a focus on determining the feasibility of use and performance of the tool. During the course of the pilot, at least two international experts with experience in piloting WHODAS will conduct periodic visits, and local experts will be hired by the World Bank to monitor the day-to-day operation of the pilot.

For fairness and transparency, the official determination of disability or needs assessment status under the relevant legislation and regulations cannot be altered during the pilot period. That is, the results of the WHODAS-36 clinical interview will not in any way affect disability or individual needs assessment determination.

Timeline: Local experts to conduct WHODAS pilot will be identified by November 22 and will be hired by the World Bank by December 22, 2021. Technical working group should be established by December 15, 20021 (Detail plan for the pilot preparation/implementation is presented in Annex 1).

II. Agreement on parameters and procedures for the pilot

The Pilot Study Working Group, in consultation with relevant stakeholders should reach consensus on the following pilot parameters and procedures:

a) Pilot time period, place of pilot and number pilot cases (at least 2,000- cases; preferably 2,500-3,000)

b) Pilot Inclusion criteria:

i. Persons in lawful age including those who have reached the age of 16.
ii. The person should have been certified for disability or by TEMC the first time (disability status has not been determined before) or repeatedly (the person seeks re-certification when the previous disability period has expired) and is applying for the needs assessment with the SAA.
iii. The cognitive capacity and language skills of the person allow to freely participate in the interview and answer to questions asked by interviewer. For persons with mental disabilities, it is acceptable that a trustee is present during the interview to help explain the content of the questions in easy language.
iv. Person or person's next of kin has signed informed consent on participation in the Pilot Study by means of face-to-face interviews (the last one can be accepted only in case when the interviewee verbally agrees to participate in the pilot study but is physically unable to put signature). In case of telephone interviews, the person has orally given informed consent on participation in the Pilot Study, which is recorded in the interview questionnaire.
v. The interview can be conducted at the SAA premises or through a visit to the applicant's residence (the preferred option to be confirmed with the Ministry). If COVID rules are still in place, the interview will be conducted over the phone (ideally in the video call).

c) Participant recruitment procedures:

i. Persons who apply (submit an application and other required documents) for the individual needs assessment will be invited to participate in the interview which will take place after they have completed and submitted to SAA the self-assessment questionnaire. It is advisable that the WHODAS interview takes place not later than a month after the disability assessment certificate has been issued to a person.

ii. Persons who submit an application for the needs assessment by regular post or in the form of an electronically signed document (i.e., the person does not appear in person to submit documents) will be invited to participate in the interview by telephone.

iii. The recruitment process should flow as follows: (i) a person who has obtained a disability certificate from TMEC submits an application for the individual needs assessment to a local SAA branch; (ii) the person receives and is instructed to fill in the currently used self-assessment questionnaire; (iii) the person fills in the self-assessment questionnaire and submits it to SAA; (iv) the SAA officer visits the household to conduct regular verification process for the needs assessment for cases when personal assistance is requested, or invites a person to an interview in the SAA premises if other benefits, services are applied for. In addition, during the visit, the SAA officer administers WHODAS 36; (v) SAA completes the needs assessment and issues a decision. As noted, the WHODAS interview

should take place not later than a month after the needs assessment application has been submitted, if possible. As noted, the interviews will be conducted over the phone (preferably via a video call), as long as the COVID rules do not allow face to face interactions.

d) Pilot instrument:

i. The 36-item, clinically administered version of WHODAS-36 will be piloted. 'Clinical administration' entails that i) all interviewers will have adequate experience to administer functioning tools, such social workers with professional expertise and competencies in the field; ii) all interviewers will be specifically trained on WHODAS; and iii) the data collected from the questions in WHODAS will be the result of a clinical determination, by the interviewer, of the correct response option to be selected, based on what is said in the interview (i.e. the interviewer may need to determine that the objectively correct response is different from that which the applicant is providing. In order to reach such a conclusion, it is likely that the interviewer will use follow-up or probing questions).

ii. Although for the purposes of analysis, the complete WHODAS needs to be administered for each pilot case, if the study team feels it is appropriate, additional questions, in the WHODAS format and using the same response options can be added in order, e.g., to include ICF categories other than those that are already included in the WHODAS. It is recommended that these additional questions should be agreed upon by all stakeholders. To avoid making additional burden on interviewees and other practical considerations, the number of additional questions should be limited to 4-5.

iii. The study team may also wish to add some basic qualitative questions at the end of the interview as a matter of debriefing in selected cases (e.g., questions such as, *Did you find the questions meaningful and was the interview conducted in a respectful manner*?).

e) Consent

Before the WHODAS is administered, the applicant should be informed about (i) the nature of the pilot and its purpose; and (ii) that they (applicants) may notice some repetition of questions, but not to be concerned and to answer honestly. **The applicants should be clearly informed that the pilot is for the purpose of improving the needs assessment system and that the pilot will have no impact on their eligibility to benefits.** A formal consent form should be developed explicitly for the pilot. The consent should follow relevant Bulgarian rules.

f) Additional preliminary parameters

i. **Interview scheduling**: Ideally, the WHODAS interview will take place during the home visit in the case of applicants for personal assistance and in an interview at the SAA premises for other applications. However, after consultations with SAA, other options may be considered (COVID related rules).

ii. **Conduct of interview**: Only qualified and specifically trained SAA's staff currently conducting individual needs assessment procedures will conduct the WHODAS interviewers. Given that the clinical interview should last between 45 minutes and an hour, a decision must be made about the number of interviews that can be conduct per day for each interviewer.

iii. **Translation:** There is an official Bulgarian translation of the ICF and the Guidelines for the implementation of the WHO Methodology for Health and Disability Assessment (WHODAS 2.0) (the output of the implemented activity under procedure BG05M9OP001-3.010 Expertise of the under OP HRD 2014-2020 with the beneficiary MLSP, and partners MH and NSSI.)

iv. **Data collection form**: In addition to the 36 questions of the WHODAS, what additional questions are added to tailor the questionnaire for Bulgaria, further information should be collected: unique identification number (if applicable), standard social-demographic questions such as age, sex, marital status, employment status, and others to be determined. Together these questions should be put into a standardized format for data entry purposes, whether manually or electronically.

v. **Qualitative data**: To complement quantitative data, a series of interviews pertaining to the needs assessment will be conducted with SAA staff, pilot participants, other stakeholders.

vi. **Data base for analysis**: Ideally the data base will contain the following information: (i) information contained in the disability/ certificate (medical diagnosis, ICD code, percentage of disability, duration of certificate and any other information contained in the certificate); (ii) data collected through a self-assessment; (iii) data collected through WHODAS-36 interview; (iv) data collected during the home visit as part of the current needs assessment process; (v) needs assessment decision. A standardized data entry format should be developed and deployed. To protect privacy, only anonymized data will be used. The data base compiled in this way will enable a comprehensive quantitative analysis. The information that is currently collected and compiled in an electronic format will be integrated into this data base.

vii. **Training**: The recruited interviewers, as well as local experts will need to be trained by WB. The one-day training session will consist of an introduction to functioning for disability, the WHODAS questionnaires, and conducting a WHODAS interview. Depending on the number of interviewers, training will be organized in batches, or training of trainers will be organized first.

Timeline: WHODAS training for social workers who will conduct WHODAS pilot will take place on December 1, 2021, and a follow up meeting to discuss any additional questions will be organized by mid-December if needed be.

Process of the Pilot Study

Subject to decisions made during the Pilot Study Technical Working Group concerning parameters and specific logistical issues that affect the timing, location or other characteristic of the study, the process of the Pilot Study should proceed as follows:

1. For the Pilot, trained interviewers will conduct interviews of applicants who meet the inclusion criteria and have been recruited using WHODAS questionnaire concurrently with the current needs assessment process described in the previous paragraph. **The results of the pilot study will not in any way affect the official procedures of comprehensive disability needs assessment.**

   Study interviewers will use WHODAS translation approved by the World Health Organization. The Questionnaire includes socio-demographic data (gender, age, etc.) and 36 questions on difficulties that the person experiences due to her/his health conditions and functional

limitations. Paper copies of the WHODAS 2.0 questionnaire for interviewing people will be provided.

Study interviewers will follow standardized procedure of the interview, as described in the WHODAS manual.[24] Before the beginning of the Pilot Study, all designated interviewers (SAA staff currently conducting individual needs assessment) will participate in one-day long course regarding WHODAS questionnaire to ensure standardization on data collection.

2. The interviews must be conducted in a way that ensures the protection of participant's personal privacy, confidentiality, as well as to avoid distraction of the participant. Interview will be performed once with one participant and up to 45 minutes of time are planned for the interview.

   During the interview, the interviewers will tick the answers of the respondent in WHODAS paper-based form, and the respondent will certify with his / her signature the correctness of the answers indicated therein.

   Afterwards all data will be entered into the previously established WHODAS input module, after which the completed WHODAS paper-based forms will be handed over to previously appointed SAA employee, who will ensure its safe storage, ensuring the protection of personal data. After the transfer of anonymized data to the World Bank researchers for analysis, the WHODAS paper-based forms will be stored for 3 months. After this period, all data (filled paper-based forms and records in the Disability Information System) will be permanently deleted (note: this specifically pertains to the WHODAS data).

3. All information obtained during the interviews shall be confidential and shall remain confidential without time limit. All persons involved in the pilot must adhere to this requirement with no exceptions and time limit.

4. As noted previously, a complete data set, inclusive of the WHODAS data (see above) will be anonymized and shared with the WB designated staff statistician for analyses.

   A data report in anonymized format will be created: Each person will initially be assigned a unique non-personal identifier, generated by the system. The data set will not include personally identifiable data, such as personal identification number, name, surname, place of residence, contact information. The data set will include personal data from the Disability Information System: gender, age, disease code (ICD-10), data on the severity of health disorders, etc. from the Medical Expertise Decision.

   Regarding WHODAS data, the data set will include individual personal data from the 36 questions of the WHODAS questionnaire (and additional n questions as decided previously), indicating the person's socio-demographic data and the answers provided. See Annex II for WHODAS Data Set.

   As errors may be detected in the data analysis process, the planned data retention period is 3 months after the transfer of the anonymized data set to the researchers for data analysis. If an extension is needed, it will be agreed on with MLSP.

[24] Measuring Health and Disability. Manual for WHO Disability Assessment Schedule WHODAS 2.0. 2010, Geneva, Switzerland: World Health Organization.

Timeline for implementation: WHODAS pilot implementation will start on December 22, 2021, ad will run until March 15. WHODAS analysis will be conducted from March 15 to May 15, 2022.

2. ANALYSIS PLAN

Tentatively, the analysis plan for the WHODAS pilot study will include at least the following analytic steps and processes:

A. ICF content comparative analysis of WHODAS and the functioning information collected for the purpose of the needs assessment.

The aim of this concept mapping exercise is to determine the overlap in ICF domain and category coverage between the current system and WHODAS and to assess whether information on functioning as collected is reliable and valid.

B. Quantitative comparison between Medical Expertise and total WHODAS Score.

The aim in the analysis is to align average, mean and standard deviation values of medical expertise with the total score of WHODAS for the pilot participants, as a preliminary step to quantitively compare the two results. Distribution curves for each set of results will be created. This is a side benefit of the WHODAS pilot and will provide evidence-based recommendations on how to include functioning adequately not only into needs assessment, but also disability assessment.

C. Descriptive Statistics of Pilot Sample.

This standard analysis will help to characterize the pilot sample for further analysis and discussion. The descriptive parameters will be standard socio-demographic variables (including work and education status) as well as prevalence of primary reported health conditions by ICD-10 Chapter. (There are existing population norms for WHODAS that can be used as comparison between general population values for the ICF categories in WHODAS and those used, by health condition, in MP-C. There are also data on correlations between WHODAS individual scores and health condition categories from previous studies that may also be compared.)

D. Psychometric Analysis: Validity and Reliability

As in previous studies of this nature, we know that a dataset of 2000 applications of WHODAS is statistically sufficient to construct Rasch models that can test reliability and construct validity, as an initial step, and to create a metric or linear scale from the data that makes it possible to directly compare WHODAS scores with TEMC decision outcomes for the same population. This is important to inform not only disability, but also needs assessment pointing to the areas of focus for assistance to optimize functioning.

E. Comparison between WHODAS and functioning information collected through needs assessment.

Timeline for implementation: Adjustment on the tool based on WHODAS results will be conducted in April; and then in early May through June needs assessment pilot testing will be conducted. Needs assessment tool revision and final draft is expected to be completed by August 15.

Timetable for the pilots' implementation

WHDAS pilot implementation	Activity	Timeline
	Interviewers, coordinators, IT experts identified	By November 22, 2021
	Interviewers, coordinators, IT expert contracted	By December 22, 2021
	WHODAS training	December 1, 2021
	Establishment of the Technical Working Group	By December 15, 2021
	WHODAS pilot	From December 22, 2021, to March 15, 2022
	Draft WHODAS Analysis	March 15, 2022- May 15, 2022