General discussion
CHAPTER 7

The experience of physical symptoms as a common event was highlighted by our studies. We found that 80% of the adult participants in the Lifelines cohort reported at least one physical symptom in the past 7 days (Chapter 3). Around 80% of young adults reported having at least 1 Functional Somatic Symptom (FSS) in the past six months (Chapter 4), and around 50% of adults visited a healthcare professional for FSS in the last year (Chapter 5). All these population-based cohort studies showed how frequent and common it is to have (un)explained physical symptoms in the general population.

The overall aim of this thesis was to address three methodological issues regarding the measurement of FSS: 1) the heterogeneity of FSS and the problems of using sum-scores, 2) the uncertainty concerning which symptoms are most relevant for FSS assessment, and 3) the ambiguity of the use of sum-scores for different subgroups. With these aims, we strove to contribute to the research on the heterogeneity of FSS and to contribute to the improvement of measurement tools assessing FSS.

In this chapter, I will discuss the main results obtained in this dissertation, followed by the methodological considerations derived from our results. Next, I will discuss the clinical and research implications of our results. Finally, recommendations for employment and construction of FSS questionnaires will be presented.

Main findings in context

Consistent with previous literature (Brown & Moskowitz, 1997; Campo, 2012; Campo et al., 2004; Henningsen et al., 2003; Janssens et al., 2010; Jellesma et al., 2006), we found that Negative Affect (NA) predicted current and future FSS (Chapters 2 and 4). However, regarding our question about the relationship between Positive Affect (PA) and FSS, we found that having a high level of PA significantly predicted current lower levels of FSS even when adjusting for covariates. Yet, the association between PA and lower levels of FSS was not statistically significant when it comes to predicting changes in FSS over time. Furthermore, even if there was a significant association between PA and current FSS, the effect of this association was small and therefore of questionable clinical relevance. Nevertheless, it is important to highlight that small effect sizes are typically interpreted as unimportant, but this ignores the fact that these estimated effect sizes are averaged over the whole group. That is, effect sizes estimated in large cohort studies do not consider the heterogeneity of FSS patients and therefore ignore the fact that a part of the individuals may have larger or smaller effect sizes than what is estimated for the group. For this reason, heterogeneity of FSS needs to be considered in FSS measurement.
Below, I will discuss the results of the studies included in our dissertation, and how they relate to the methodological gaps identified in the introduction.

1) **Heterogeneity of FSS and the problematic use of sum-scores**

In clinical settings, IRT aims to measure the latent severity of a clinical construct in participants of a study (Reise & Waller, 2009). Under the IRT framework, participants are characterized on a latent FSS severity continuum rather than receiving a sum-score based on a mere count of symptoms. IRT-based person scores are calculated by considering both the response patterns of each participant and the characteristics of the items. IRT person scores might thus provide an alternative to the methodological problem of using sum-scores, given that they provide a more accurate measurement of FSS severity and a clearer understanding of what we are measuring (i.e., FSS severity in a continuum). However, the consequences of an IRT-based approach for empirical results remained unclear.

What are the differences between using sum-scores and IRT-based person scores when studying FSS? According to what we found in Chapter 4, the correlation between sum-scores and IRT person (θ) scores of FSS was higher than 0.9. These very high correlations have been observed in previous reports (Embretson & Reise, 2013). Moreover, there were negligible differences in the epidemiological study on the relationship between PA and FSS when replacing the sum-score with the IRT person score as the FSS outcome measure. PA still significantly predicted current FSS levels with a small effect.

Based on the results from Chapter 4, IRT-based person scores may not seem to show great advantages over the use of sum-scores in an epidemiological study and seem to be more complicated to calculate. One may wonder whether it would be better to simply use sum-scores instead of IRT-based scores. This is not that simple to be judged. Although the sum-scores were highly comparable with the IRT-based person scores, the IRT-based person scores managed to capture heterogeneity between FSS patients better than the sum-scores. More specifically, in Chapter 4, persons with the same mean sum-score of 0.3 (in a scale of min = 0, max = 1.67) had IRT-based person scores that varied from 0 to 1 (in a scale of min = -0.93 to max = 3.36). This means that IRT person scores captured more heterogeneity among subjects with similar sum-scores. Additionally, as sum-scores of different FSS questionnaires often vary in the number of symptoms and answer options included, it is difficult to directly compare these sum-scores obtained from different questionnaires. IRT-based scores allow us to have a more straightforward interpretation of what we are measuring and where persons are located in the spectrum of the construct. Furthermore, IRT also provides detailed estimations of the Standard Error (SE) of
measurement along the FSS continuum, which provides important information about the accuracy of the person scores along the FSS latent trait. For example, we learned from Chapter 3 that the SCL-90 provides lower SE for participants with IRT person scores between 1 and 4, which was the top 28% of the sample.

In sum, although at first glance it may seem that replacing sum-scores with IRT person scores does not have a great added value in epidemiological studies, IRT person scores can address the heterogeneity of FSS severity more accurately than sum-scores. Furthermore, their interpretation is more straightforward given that their standardized scores locate persons on an FSS severity continuum. This improves comparability between studies using different FSS measures. Moreover, it has been suggested that IRT-based scores provide unbiased estimations when studying within and between-person variations in clinical longitudinal studies (Gorter et al., 2015), and they could be useful in the future development of Computerized Adaptive Testing (CAT), providing more accuracy in the measurement (Gibbons et al., 2016).

2) It is unknown which symptoms are most relevant for FSS assessment

Our results showed that experts' opinions and data-driven results on which symptoms to include in FSS questionnaires do not necessarily coincide. In a systematic review analyzing 40 FSS questionnaires, 70% of them included the item “headaches”, 65% included “nausea/upset stomach”, 58% included “shortness of breath”, 55% included “dizziness”, and 55% included “low back pain/backaches” (Zijlema et al., 2013). The symptoms included in these questionnaires have typically been selected based on experts' knowledge. However, some of these items were not highlighted as the most informative in the assessment of FSS severity according to our data-driven results.

These results signal some inconsistencies between empirical evidence and experts' opinion. First, the most popular item to measure FSS among all questionnaires according to experts (i.e., headaches), showed to be the least discriminative for FSS severity among adults from the general population (Chapter 3). “Headaches” showed slightly better discrimination for a population-based group of younger adults (Chapter 4) than in the cohorts including adults of all ages. However, this item did not seem to contribute much to the measurement of FSS in any of our studies. Second, “low back pain/backaches” also showed very low discrimination abilities in adults (Chapters 4 and 5), despite being one of the most popular items in FSS questionnaires. Third, some of the most discriminative items (i.e., items that best distinguish between individuals' FSS severity levels) found in our studies are not commonly included in existing FSS questionnaires. Specifically, the items “heavy feelings in arms or legs” and “feeling weak physically” of the SCL-90 questionnaire were found...
General Discussion

to be the best at discriminating between FSS severity levels in adults from the general population (Chapter 3). Similarly, “localized (muscle) weakness” was the most discriminative item from the CIDI in an adult cohort where unexplained symptoms were specifically assessed (Chapter 5). Studies with participants from psychiatric (Paap et al., 2011) and neuromuscular (Hart et al., 2012) clinical populations have found similar results, especially regarding items related to “weakness”. These results indicate that “weakness” is a discriminating symptom across questionnaires and populations.

We also found one item highlighted as relevant by both our analysis and experts’ opinion. In Chapters 4 and 5, we found that the item “nausea” was one of the most discriminating symptoms, especially for the group of younger adults. This is consistent with experts’ opinions since nausea is one of the most frequently included symptoms in FSS questionnaires. From our results, we observed that there is a possibility that nausea can be more discriminating when it is assessed as a medically unexplained symptom (Chapters 4 and 5) than when it is assessed as a general symptom (Chapter 3).

A possible explanation for the discrepancies between experts’ opinion and data-driven results could be that symptoms such as “headache” and “low back pain” tend to be commonly observed and could be explained by several common reasons (e.g., tension headache, muscle strain, poor posture, etc). This indicates that persons with FSS, as well as great part of the population, will report these symptoms and therefore these may not be able to differentiate among individuals with different FSS severities. On the other hand, feeling weak in muscles or parts of the body or heaviness in extremities are not very common symptoms and do not have many common explanations.

According to our findings, some items were found to be less relevant for FSS assessment. Two of our studies (Chapters 4 and 5) showed that items related to vision problems, namely “double vision”, “blurred vision”, and “eye problems”, reflected the highest FSS severity. However, they showed low discrimination abilities and low factor loadings in the factor analyses and principal component analyses. The item “rashes or skin problems” showed the poorest discrimination and very low factor loadings in the Adult Self Report (ASR) questionnaire administered to the young adolescent cohort (Chapters 2 and 4). In Chapter 6, these two items also showed difficulties at distinguishing between latent classes of FSS. Low factor loadings have also been found for these symptoms in previous studies with younger participants (Janssens et al., 2011, 2014). These results suggest that items related to skin and eye problems may not be very informative for measuring FSS severity.

In sum, we found some discrepancies between the most often included symptoms in the expert-based FSS questionnaires and the data-driven results. According
to our results, symptoms related to "weakness", "heaviness in extremities", and "nausea" are most discriminative of FSS severity. This makes them more relevant and accurate for the assessment of FSS severity than symptoms that are commonly included in FSS questionnaires, such as "headache" or "back pain".

3) **Sum-scores may not have the same meaning for different subgroups**

FSS can have very heterogeneous patterns of symptom manifestation, and several factors can influence the reporting of symptoms (i.e., sociodemographic factors, comorbidity, etc.). In Chapter 5, we further examined this problem by studying Differential Item Functioning (DIF) by sex and age. The goal of this study was to investigate whether the probability of reporting a symptom could be different for men and women, and for young to middle-aged adults (< 60 years of age) compared to older adults (60+ years of age), even when they presented the same FSS severity level. In other words, we wanted to explore if items are biased towards a group in the reporting of FSS.

According to the results from Chapter 5, the item “abdominal pain” was identified to have reporting bias based on sex. This item was slightly more discriminative for women than for men. In addition, women tend to report this item when having lower levels of FSS severity, while men report it when having higher levels of FSS severity than women. This is in line with previous studies finding that functional abdominal pain and gastrointestinal symptoms tend to be more frequently reported by women (Ballering et al., 2020; Korterink et al., 2015; Rajindrajith et al., 2018). Interestingly, “abdominal pain” was shown to be very discriminating of FSS severity according to the results from Chapter 4. This suggests that this item is relevant for the measurement of FSS severity, but sex should be considered as a factor that can influence the reporting of the symptom. The item “loss of touch or pain sensation” also showed reporting bias. We found that when women report this item, they may be showing a higher FSS severity level overall compared to men. It was also shown that this item seems to be good at discriminating FSS severity levels in men, but not so good at discriminating in women. Previous studies suggest that the experience of pain perception differs by sex, and that women’s perception of pain may fluctuate with hormonal changes (Sun et al., 2019). This could be one of the reasons why “loss of touch or pain sensation” may not be very accurate at discriminating between FSS severity levels among women.

Regarding age, there was no consistent evidence of reporting bias. However, we found that “joint pain” was not a discriminative item of FSS severity in older adults (60+ years) compared with younger to middle-aged adults. This could be explained by the higher likelihood of joint pain and other musculoskeletal problems arising from pathology at older ages. That is, older adults and their
physicians would be less likely to attribute joint pain to medically unexplained reasons. Other common symptoms at older ages, such as chest pain and shortness of breath (Sha et al., 2005), showed small indications of reporting bias. This implies that these common symptoms at an older age may be less accurate for the assessment of FSS severity in older patients than in younger patients.

Finally, in Chapter 6, we explored the heterogeneity of FSS with respect to latent subgroups. We intended to explore the existence of subgroups of individuals with FSS based on symptoms and to explore differences in symptom severity and discrimination among these subgroups. We found very little evidence for the existence of subgroups based on symptom patterns (e.g., gastrointestinal, musculoskeletal, etc.), but we did find an indication of subgrouping based on the severity of symptoms. That is, symptom severities were found to be more relevant than reported symptom patterns to differentiate between underlying subgroups. This implies that IRT parameters essentially remained invariant across the sample, i.e., the items mean the same across groups. Thus, as previous studies have also highlighted (Fink et al., 2007; Rosmalen et al., 2011), FSS latent classes may not arise from different combinations of symptoms with varying severities, but severity alone could be the most relevant factor to classify groups of patients.

Methodological considerations

Several methodological aspects should be considered in the interpretation of these results, as well as for future studies on FSS.

Firstly, across all our studies, the data was zero-inflated (ZI). This means that the data distributions were very skewed to the right. That is, a large part of the sample did not report any symptoms, leading to many zeros in the data (Loeys et al., 2012). This was reflected, for example, in the distribution of the answer options from the analysed questionnaires. In both the FSS scales of the SCL-90 and the ASR, on average, between 75% and 78% of the answers to the question of experiencing a symptom were “Never/Not at all (0)”. Using the FSS scale from the CIDI, around 95% of the responses, on average, signalled the absence of an FSS symptom. On the contrary, reporting the presence of a symptom or reporting the experience of a symptom “extremely” or “often or a lot” concerned less than 5% of the answers on average in all FSS questionnaires (Chapters 3, 4, and 5). These types of data distributions are to be expected when assessing clinical constructs (Reise & Waller, 2009), especially in non-clinical populations. Yet, this can lead to floor effects and interference with the normality assumption which is often assumed in the estimation of statistical
models. To deal with zero-inflated data, we performed sensitivity analyses to explore the impact of this data distribution on the main results. We applied a Zero Inflated (ZI) IRT model (Wall et al., 2015) in Chapter 5, while we excluded participants with sum-scores of zero in Chapter 6. In both cases, these sensitivity analyses did not show a large impact on the results.

Secondly, although a big percentage of the participants reported one or more symptoms in all studied samples, few of them presented a high level of FSS severity. This influenced the estimation of IRT parameters in Chapters 3, 4, and 5. Because of this, we observed that only participants with a very high level of FSS (i.e., more than two SD from the mean) reported the presence of a symptom or experienced a symptom “often” or “a lot”. This is typically observed in clinical instruments applied to non-clinical populations (Reise & Waller, 2009). It is likely that when applying IRT in a clinical FSS population, participants would have a higher likelihood to report symptoms, and because of this, the estimates of FSS severity level could appear lower when directly compared with non-clinical populations. This would not mean, however, that clinical populations possess lower FSS severity levels, but that because of the sample, the magnitude of the IRT parameters would be lower in clinical FSS populations than in non-clinical populations. This aspect should be considered when directly comparing results obtained in such populations to our studies in general population-based cohorts. Future studies could explore the parameters of items in populations with a high prevalence of FSS, to explore potential differences in the estimation of item parameters.

Thirdly, the assumption of unidimensionality of FSS was broadly met in our studies, according to the heuristic indication that at least 20% of the total variance should be explained by one factor (Reckase, 1979). However, there are some aspects to consider regarding the dimensionality in our studies. In all our studies, a one-factor solution explained less than 35% of the total variance. Additionally, a two-factor solution was found to be appropriate in exploratory factor analyses of the ASR questionnaire in the young adolescent cohort (Chapters 2, 5, and 6). This solution consisted of a dimension of general physical symptoms, and a dimension of headache and gastrointestinal symptoms. On the other hand, when performing the latent class analysis (Chapter 6), we did not find that participants clustered according to different types of symptoms. In the past, clusters of FSS symptoms (e.g., musculoskeletal, gastrointestinal, cardiopulmonary) have been suggested and identified in factor and cluster analyses (Fink et al., 2007; Gara et al., 1998; Lacourt et al., 2013; Rosmalen et al., 2011; Tsai, 2010), while evidence has been found of a higher-order FSS factor (Fink et al., 2007; Rosmalen et al., 2011). Our results evoke the “lumpers vs splitters” discussion (Wessely et al., 1999) in which some FSS researchers advocate for the discrimination of different patterns of symptoms (splitters), while others argue that all symptoms represent one underlying construct (lumpers). Our results do
General Discussion

not provide a straightforward answer related to this discussion but results from Chapter 6 suggest that the overall severity of FSS could be more relevant for FSS assessment than the type of symptoms.

Finally, we relied on self-reported measures of symptoms in all our studies. This means that symptom reporting could have been affected by recall bias (e.g., in the CIDI questionnaire asking for the presence of symptoms in the last 12 months), or that participants could have reported symptoms with an underlying medical explanation that we were not aware of. To deal with the latter issue, in Chapter 3, we excluded participants based on their self-report of somatic conditions (e.g., diabetes, asthma, COPD) from the main analysis. However, when performing sensitivity analysis including participants with and without somatic conditions, the results were comparable to the main analysis. In Chapter 5, the CIDI questionnaire employed an algorithm to determine if symptoms were explained or unexplained, and although medical conditions were self-reported, the algorithm was based on specific questions related to the medical trajectory leading to the diagnosis. This aimed to ensure that the symptoms were adequately classified as unexplained.

Clinical and research implications

These studies highlighted three main clinical implications. First, discrepancies were found between the symptoms that are most frequently included in FSS questionnaires based on experts’ knowledge and those that are most relevant according to empirical data. This highlights the relevance of studying the properties of individual items and their contribution to the overall measurement of FSS. This does not mean, however, that clinicians should stop assessing or should dismiss other symptoms when patients report one of the most discriminative or severe FSS symptoms. For clinicians, these symptoms could be useful to target patients that could need an augmented assessment of FSS. Second, it is relevant for clinicians to consider the sex and age of the patients when inquiring about certain FSS symptoms such as abdominal pain, loss of touch/pain sensation, or joint pain. Our findings underlined that these items may provide biased assessments of FSS, depending on the sex or age of the individuals. For instance, it would not be very useful to assess FSS levels with the item “joint pain” in older adults, since this item is not discriminative for this age group. Thus, clinicians should be careful when assessing and interpreting the reporting of these symptoms in different sex and age groups. Third, we did not find subgroups of participants based on their symptoms. Instead, the overall severity of FSS in all symptoms is indicated to be the most relevant for subgrouping. Clinicians could take this into account for screening in primary care, paying extra attention to the severity of the symptoms rather than at
Recommendations for instrument use in FSS

As for the implications of these findings for research, the results from this study could greatly benefit the construction and refinement of FSS questionnaires. First, the most discriminative items found in these studies could be considered for inclusion in FSS questionnaires, in contrast to items that hinder the accuracy of measurement (e.g., skin problems and eye problems). Second, the person scores derived from IRT models, although highly correlated with sum-scores, provided more insight on the heterogeneity of the sample, more accuracy on the measurement, and a straightforward and standardized interpretation. Additionally, IRT-based scores could be more accurate than sum-scores for studying between-person variations in longitudinal studies (Gorter et al., 2015). Using IRT-based scores could therefore enrich research and provide robustness to the results on FSS studies. Third, constructing more informative questionnaires based on an IRT approach could facilitate the comparison between FSS studies, which is a major obstacle in the field. The results of these studies can contribute to the construction of shorter and more accurate questionnaires that will be less burdensome for patients and more efficient for clinicians and researchers to use.

Recommendations for instrument use in FSS

Applying IRT to FSS instruments consistently underlined that some items could be more relevant for the measurement of FSS than others. Our studies also addressed the heterogeneity of FSS severity by exploring IRT-based person scores, and the influence of subgroup belonging on the reporting of specific items. As previously mentioned, these results could contribute to the construction of new shorter, more reliable, and more accurate assessment scales. Nevertheless, the question of how to choose a questionnaire from the more than 40 questionnaires that are already available remains open.

We explored the frequency of the items used in the 40 FSS questionnaires identified by Zijlema et al. (2003) with the aim to determine which of the existing questionnaires could be the most useful at measuring FSS according to our results. Table 1 shows a summary of the properties, advantages, and disadvantages of the identified questionnaires. As expected, questionnaires with the largest number of items, such as the Screening for Somatoform Symptoms-7 (SOMS-7) with 53 items (Rief & Hiller, 2003), or the Somatic Symptom Index (SSI) with 35 items (Escobar et al., 1989), included most of the items that we identified as most relevant for FSS measurement. Both aimed to assess the construct of somatization (which assumes that symptoms are unexplained). However, besides their high number of items, the assessment time frames that they use (past seven days, and lifetime, respectively), may not be the most suitable for combinations of qualitatively different symptoms (e.g., only gastrointestinal).
measuring FSS (Joustra et al., 2018; Leiknes et al., 2006).

According to a previous study, a 4-week time frame, followed by a 3-month time frame, reflects clinically relevant somatic symptom severity best (Joustra et al., 2018). We found a 25-item scale with a measurement time frame of 4 weeks, that included most of the relevant items to measure FSS while excluding the items that presented problems in the assessment of FSS according to our results (e.g., skin and eye problems). This was the Bodily Distress Scale (BDS, (Budtz-Lilly et al., 2015)) which is a scale including four dimensions of symptoms: Cardiopulmonary/autonomic (arousal), gastrointestinal symptoms, musculoskeletal symptoms, and general symptoms. This questionnaire is not restricted to medically unexplained symptoms but asks whether participants have been bothered by these symptoms in the last four weeks. It included four items that we identified as most relevant: “nausea” and “abdominal pain” from the gastrointestinal symptoms dimension, and “localized weakness” and “pain in arms or legs” from the musculoskeletal dimension. Although this scale seems to be promising as an FSS measure, it still presents the problem of being relatively lengthy and potentially burdensome for participants.

We did find a short questionnaire that included all the important items identified by our studies, and that excluded items that were not relevant for measurement, namely, the somatization disorder screening index developed by Swartz et al. (1986). This is an 11-item instrument constructed based on the somatization section of the Diagnostic Interview Schedule (DIS). Swartz et al. (1986), examined the most frequently endorsed symptoms of the DIS among respondents diagnosed with the DSM-III somatization disorder and selected 11 medically unexplained symptoms which discriminated between respondents with and without somatization disorder, based on a frequency of selection greater or equal to 50%. Additionally, they tested the ability of these 11 symptoms to predict somatization disorder. In the end, the index was able to identify 97.6% of the participants with somatization disorder, and 99% of the participants without the disorder, in community samples (Swartz et al., 1986). It included the items “weakness”, “nausea”, “abdominal pain”, and “pain in extremities”. Additionally, this index excluded symptoms that are not as informative as others, such as headache, back pain, skin, and eye problems. However, one major problem of this scale is that it assesses symptoms in a lifetime timeframe, which has proven to be a very unreliable timeframe to assess FSS (Leiknes et al., 2006). Therefore, it is not advisable to use this scale in its current state. It would be worthwhile to conduct validation studies for this scale but assessing FSS on a time frame of 4 weeks.

A shorter questionnaire that included the items “heaviness in arms/legs”, “feeling weakness in individual body parts”, and “stomach pains”, was the 7-item somatoform complaints scale of the HEALTH-49 questionnaire (Rabung
et al., 2009). Although this scale assesses in a time frame of two weeks, which has also shown not to be optimal compared to 4-weeks (Joustra et al., 2018), it could be further validated since besides being short, it included important symptoms and excluded less relevant symptoms.

A remark about the item “abdominal pain” is that, even though it showed to be discriminative in Chapter 4, sex should be considered when assessing this item. In Chapter 5, this item consistently showed reporting bias, i.e., men tend to report it when having higher levels of FSS severity than women. Clinicians and researchers should take this into account when including this item in FSS assessment.

Zijlema’s et al. (2013) review recommends the use of the somatization subscale of the SCL-90 (Derogatis et al., 1974), and the Patient Health Questionnaire 15 (PHQ-15; Kroenke et al., 1998) for large scale studies of common somatic symptoms. This is because these scales have been extensively validated, contained the most relevant symptoms according to experts, were short in length, and enquired about symptoms in the last month or week, and they were not restricted to medically unexplained symptoms, but symptoms in general. Given their extended use and evidence of validity, these questionnaires are valuable for the measurement of common somatic symptoms, as well as FSS. One disadvantage of the PHQ-15 is that some of its items are not very precise regarding the symptoms that are being measured, such as “nausea, gas or indigestion”, and “pain in your arms, legs, or joints”. This could affect how discriminative the items are. It would be worthwhile to further examine the properties of the PHQ-15 items with IRT since this questionnaire has an adequate measurement timeframe (4 weeks) and does not include symptoms such as skin and eye problems in the measurement. The SCL-90 on the other hand defines the symptoms more precisely, although the measurement timeframe is not optimal (past 7 days).

An additional point for discussion is whether questionnaires should ask for the presence of symptoms regardless of them being unexplained. In Chapter 3 we saw that there were no major differences in the properties of the items when assessing participants with and without somatic conditions in the general population. Moreover, there were consistencies regarding the relevance of some items between the questionnaires assessed, regardless of them questioning whether symptoms were explained or unexplained (e.g., the SCL-90 inquired about symptoms in general). Although the core definition of FSS is that these are unexplained by a medical pathology, it is certainly difficult to determine if symptoms are unexplained, and even more so, on the basis of self-report. This is something to consider in the construction of future questionnaires. Overall, the improvement of FSS questionnaires is yet to be achieved. However, the results described in this thesis and previous studies (Joustra et al., 2018; van
General Discussion

Driel et al., 2018; Zijlema et al., 2013) have provided valuable insights about the most relevant symptoms for measurement of FSS and the most adequate assessment time frames.

Concluding remarks

This thesis addressed some of the methodological gaps regarding FSS measurements by means of IRT methods. The studies indicated which symptoms are most relevant for the assessment of FSS severity, and the reporting of which symptoms is likely to be biased by age and sex. Additionally, although IRT-based person scores are highly correlated to sum-scores, they provide more variability which could be useful in within-subject longitudinal studies addressing the problem of heterogeneity. An important finding was also that the overall severity of FSS could be more important for identifying and classifying individuals than the combination of symptoms that they may present. IRT has important features to offer in clinical research, improving the accuracy of measurement, and providing the basis for personalized assessment from a data-driven perspective that can benefit both research and clinical practice.
Table 1. Summary of the suitable questionnaires identified for the measurement of FSS

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Number of Items</th>
<th>Assessment Time Frame</th>
<th>Most Relevant Items Included</th>
<th>Type of Symptoms</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
</table>
| Screening for Somatoform Symptoms-7 (SOMS-7; Rief & Hilker, 2003) | 53              | Past seven days       | Abdominal pain, Nausea, Paralysis/localized weakness | Unexplained symptoms | Inclusion of some of the most relevant items.                                                 | - Large length
|                                                    |                 |                       |                             |                      | Assessment time frame.                                                                       | - Inclusion of symptoms that were not informative enough/presented problems in assessment (headaches, back pain, diarrhea, blindness, double vision, skin discoloration) |
| Somatic Symptoms Index (SSI; Escobar et al., 1989) | 35              | Lifetime              | Muscle weakness/paralysis, Abdominal pain, Nausea | Unexplained symptoms | Inclusion of some of the most relevant items.                                                 | - Large length
|                                                    |                 |                       |                             |                      | Assessment time frame.                                                                       | - Inclusion of symptoms that were not informative enough/presented problems in assessment (back pain, diarrhea, blurred vision, blindness, double vision) |
| Bodily distress Scale (BDS; Budz/Lilly et al., 2015) | 25              | Past four weeks       | Nausea, Abdominal pain, Localized weakness, Pain in arms or legs | Symptoms irrespective of their cause | Assessment time frame.                                                                       | - Large length
|                                                    |                 |                       |                             |                      | Inclusion of most relevant items.                                                             | - Exclusion of items that presented problems in assessment (skin and eye problems) |
| Somatization disorder screening index (Swartz et al., 1986) | 11              | Lifetime              | Weakness, Nausea, Abdominal pain, Pain in extremities | Unexplained symptoms | Short length                                                                                  | - Assessment time frame                                                                 |
|                                                    |                 |                       |                             |                      | Inclusion of most relevant items.                                                             | - Exclusion of items that were not informative enough/presented problems in assessment (headache, back pain, skin, and eye problems) |
| Somatoform complaints scale of the HEALTH-49 (Rabung et al., 2009) | 7               | Past two weeks        | Heaviness in arms/legs, Feeling weakness in individual body parts, Stomach pains | Unexplained symptoms | Short length                                                                                  | - Assessment time frame
|                                                    |                 |                       |                             |                      | Inclusion of most relevant items.                                                             | - Inclusion of symptoms that were not informative enough (headaches, back pain) |


<table>
<thead>
<tr>
<th>Instrument</th>
<th>Items</th>
<th>Symptoms Irrespective of Their Cause</th>
<th>Assessment Time Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>Somatization subscale of the Symptom Check List-90 (SCL-90; Derogatis et al., 1974)</td>
<td>- Heaviness in arms/legs&lt;br&gt;- Nausea or upset stomach&lt;br&gt;- Feeling weak in parts of your body</td>
<td>- Short length&lt;br&gt;- Inclusion of some of the most relevant items.&lt;br&gt;- Exclusion of items that presented problems in assessment (skin and eye problems)&lt;br&gt;- Identified as the most suitable for use in large scale studies of common somatic symptoms (Zijlma et al., 2013)</td>
<td>- Inclusion of symptoms that were not informative enough (headaches, pains in lower back)</td>
</tr>
<tr>
<td>Patient Health Questionnaire 15 (PHQ-15; Kroenke et al., 1998)</td>
<td>- Pain in stomach&lt;br&gt;- Nausea, gas, indigestion</td>
<td>- Assessment time frame.&lt;br&gt;- Short length&lt;br&gt;- Inclusion of some of the most relevant items.&lt;br&gt;- Exclusion of items that presented problems in assessment (skin and eye problems)&lt;br&gt;- Identified as the most suitable for use in large scale studies of common somatic symptoms (Zijlma et al., 2013)</td>
<td>- Unprecise item wording regarding which specific symptom is being measured.&lt;br&gt;- Inclusion of symptoms that were not Informative enough (headaches, back pain, diarrhea)</td>
</tr>
</tbody>
</table>
References


CHAPTER 7


