# Promoting early recognition of persistent somatic symptoms in primary care

*Summary*

It is estimated that up to 10% of the general population experiences persistent somatic symptoms (PSS). PSS are symptoms that cannot be fully attributed to well-established biomedical pathology or to objectively determined anatomical or functional disease severity. The disease burden of PSS is often high for patients, physicians, and society. General practitioners (GPs) regularly experience difficulties in recognizing PSS and may search for a primary biomedical or psychological origin of the complaints for a long time. This approach may largely be unsuccessful due to the multidomain origin of PSS. Diagnostic difficulties are further complicated due to ambiguity in definitions and terminology of PSS. This complexity of diagnostics contributes to diagnostic delays and increases the burden of PSS, for instance by affecting the doctor-patient relationship, but also because it hampers timely and appropriate intervention. The main objective of this thesis was to promote early recognition of PSS in primary care. Due to the time constraints in primary care, the high health care utilization of patients with PSS, and the availability of large electronic medical record (EMR) datasets, the viability of a data-based approach towards predictive modeling of PSS was explored. A stepwise construction of the studies in this thesis leads to a comprehensive overview of the data available and needed to enable early data-based identification of PSS.

One of the preliminary steps towards predictive modeling of PSS was to map predictors of PSS from scientific literature. The goal of this step was to provide strong underpinning for theory-based predictive modeling using routine care data. Furthermore, this would provide insight into the availability of relevant data in GPs EMRs. Findings from the systematic review in **chapter 2** show that risk factors from the biomedical domain are currently dominant in scientific research on predictors of PSS. However, >250 predictors from all domains of the dynamic biopsychosocial model (i.e., biological, psychological, interpersonal, contextual, and health behavior) were identified. Of those, 46 were identified with adequate consistency in multiple studies. Overall, the review provides strong evidence that factors from all domains of the dynamic biopsychosocial model are important for PSS-onset.

This suggests that a broad view on all possible related factors would enhance diagnostic accuracy in PSS.

Further preliminary investigations towards predictive modeling of PSS were related to determining an PSS outcome classifier and explore the GPs needs regarding PSS. The main obstacles towards developing a clinical prediction rule for PSS based on routine primary care data is the lack of a gold standard for PSS diagnosis and a data-based outcome classifier. To get insight in registration behavior and needs of GPs, a survey was distributed amongst Dutch GPs. The results of this survey shown in **chapter 3** demonstrate that the lack of an unambiguous method of identifying PSS is not only problematic from a research perspective, but also a problem for approximately half of GPs. Results show that GPs are likely to use a variety of structured (i.e., symptom-, diagnostic-, or generic-codes) and unstructured (i.e., free text) methods to register PSS. In addition to providing insight into registration practices of GPs, results from the survey confirmed that many GPs need more support or additional tools for consultations with patients (at risk of) PSS or PSS classification.

To select the most viable method for data-based identification of PSS, four methods were derived from the survey, in combination with clinical and literature-based knowledge. In **chapter 4**, the four methods were analyzed and evaluated. Results showed that a combination of three methods would enable the most accurate data-based identification of PSS. The final identification method consist of a combination of (1) clinical coding of PSS-syndromes, (2) unstructured episode descriptions of PSS, and (3) recorded outcomes of screening questionnaires. PSS-syndromes with clinical codes in Dutch EMRs are irritable bowel syndrome (IBS), fibromyalgia (FM), and chronic fatigue syndrome (CFS). Unstructured episode descriptions can include a term synonymous to the aforementioned syndromes, but also include other PSS-syndromes (e.g., interstitial cystitis, vulvodynia, and tension headache), or be synonymous to PSS (e.g., functional complaints, somatoform, and unspecific pain). Finally, screening questionnaires include the 4-dimensional symptom questionnaire (4DSQ) which is regularly recorded in Dutch routine primary care data. A score of >20 on the somatic symptom subscale indicates PSS. With a total prevalence PSS of 8.6% in the study population, the combination of these three methods would approach general population prevalence of PSS.

Due to the knowledge that PSS has a multidomain origin and to overcome the overall predominance of biomedical structured data in routine primary care data, the viability of mental health registrations of predicting PSS diagnosis was tested in **chapter 5**. The prediction model in this chapter used mental health-related registrations from the five years directly prior to the PSS index date. In addition, it focused on the most common PSS syndromes with active ICPC codes in Dutch GP EMRs, namely IBS, FM, and CFS. This enabled further insight into similarities and differences between PSS-subtypes. The results showed that mental health registrations can predict PSS diagnosis with high accuracy. Model performance was different between PSS subtypes, with models for FM and CFS having the highest prediction value (AUC= .88 and AUC= .82, respectively) and the model for IBS being least predictive (AUC= .76). Although quantity of predictors was markedly lower for CFS, predictors generally overlap between PSS-subtypes (especially anxiety, psychosis, concentration disorders, addiction, and mental health-related referrals), while some factors may be unique to a specific syndrome, for example irritability and feeling old in IBS, delirium and developmental issues for FM, and disability due to mental illness for CFS.

Finally in **chapter 6**, results from all previous chapters are brought together to explore the optimal model for early prediction of PSS. For this study, routine care data from 76 general practices in the Netherlands were used, with an inclusion of 94,440 patients for the analyses. Candidate predictors were identified 2 to 7 years prior to PSS index date. The outcome was determined by combining the three data-based methods derived from **chapter 4**. To make optimal use of the large body of data and possibility to derive multidomain predictors, seven approaches were used to construct candidate predictors. First, two theory-driven approaches were used to extract candidate predictors. For this approach a combination of structured data was used to construct candidate predictors based on the systematic literature review in **chapter 2**. In addition, based on the knowledge derived from **chapter 3** that GPs psychosocial and behavioral indicators of PSS are most likely to be reported in the journal text, free text descriptions were extracted to form candidate predictors. Second, three non-temporal data-driven approaches were used to construct candidate predictors. Structured multidomain data from the EMRs, including symptom- and disease-codes (i.e., based ICPC-codes), medication prescriptions (i.e., based on ATC-codes), and referrals, were dichotomized. Third, utilizing machine learning

techniques, two temporal data-driven approaches were used to construct candidate predictors. The same structured multidomain data used in the non-temporal data driven approaches were used to find relevant temporal patterns. In addition, lab results were contextualized using relative grounding (i.e., indicating stable, increased, or decreased values of a particular lab test). Finally, LASSO logistic regression was applied to build 12 prediction models using a variety of combinations of 545 candidate predictors derived from the methods described above. This resulted in one baseline model, seven models based on each unique extraction approach described above, three models containing predictors from each extraction subcategory (i.e., theory-driven, and temporal and non-temporal data-driven models), and one model for which all candidate predictors were utilized (i.e., the full model). The full model showed that there is an underrepresentation of psychosocial predictors compared to what is expected based on the literature. Nonetheless, the used approaches were able to predict PSS registration with moderate certainty (AUC= .72). Despite the variety of candidate predictor extraction approaches, performance was fairly equal between the models (AUC's between .70 and .71). The performance of these models are markedly lower than the models presented in **chapter 5**. Although further research is needed to confirm this, the most notable differences are the lack of a prediction gap and the focus on specific PSS-syndromes in **chapter 5** compared to **chapter 6**.

In conclusion, this thesis provides comprehensive evidence that the multidomain nature of PSS makes the identification of PSS highly complex. The lack of an unambiguous system to diagnose and classify PSS is problematic from both a clinical as well as a research perspective, and GPs report a need for support to improve the care for patients with PSS. Despite registration difficulties, the results of this thesis show that analysis of routine primary care data can be used to develop tools to promote early recognition of PSS. Findings from this thesis indicate that the registrations of psychosocial factors should be improved to promote the reusability of data and to improve early recognition of PSS. Finally, for clinical purposes, the early prediction of PSS can be promoted based on this thesis by implementing a relatively simple non-temporal data-driven model based on ATC (medication) or ICPC (symptom and disease) codes. Such a clinical decision rule should be implemented into the GPs EMR and flag patients at risk of PSS. Flagged patients would ideally receive multidomain care using an integrative (multi-track) approach, in which there is attention for psychological, social, interpersonal, and contextual factors, in addition to keeping track of

any biomedical deterioration. In primary care this could result in (earlier) referral to interdisciplinary health care resources who may screen the patient for PSS and other multidomain problems such as mental health problems or problems in the social realm. If this results in earlier detection of PSS or other problems beyond the biomedical domain, this could enable earlier intervention which may limit deterioration of or even lead to recovery from symptoms. This will eventually result in lower health care utilization and cost, less pressure on GPs and lower disease burden for patients.