Development and Validation of the Inpatient Disruptive Behavior Index (InDiBI) for testing the effectiveness of Medical Psychiatric Units (MPU)

E. van Oostrum

S1404210

Master Thesis Clinical Psychology

Internal Supervisor: Dr. K. de Jong

Institute of Psychology

Universiteit Leiden

External Supervisor: Prof. Dr. J.J. van Busschbach, Sectie Medische Psychologie en Pyschotherapie, Erasmus MC

26-02-2019

Preface

A year ago I was given the chance to choose my own research topic for my master thesis out of an extensive list provided by the university. After reading into all topics I decided I wanted to do it differently. I was ready to go in the field, to explore something else. As I have always been interested in combining physical and mental health, I searched through a database of the Erasmus MC and I stumbled upon an interesting topic; how to evaluate the effectiveness of a new department, the Medical Psychiatric Unit, focusing on disruptive behavior. The aim was to develop an instrument that could be used in a randomized controlled trial to evaluate the effectiveness of this department. I sent an email and met up with the supervisor Jan van Busschbach and after his clarification of the project, I decided to go for it.

Along this journey, I learnt to create something out of nothing. We started with a small idea and then developed it into a whole new instrument. I cannot even remember how many articles I have read or how many rounds of discussion I have had with the other researchers who were involved. It was a long process and I would lie if I wrote down that it always was fun and that everything went smoothly. I have been frustrated, felt hopeless and I have been stressed out. However, it would not have been me if I would not have pushed through and I am glad I did. Now that I am finished I can say I am proud of the result and of myself.

However, I also have to give some credits to other people, because I could not have made it this far without them. Therefore, I would like to thank all the people who were involved in this process. Most importantly, my external supervisor Jan van Busschbach, thank you for all your time, guidance and hilarious anecdotes; it was great working with you. Reinier Timman, thank you for sharing all your statistical knowledge and learning me so many new tricks. Maarten van Schijndel, thank you for you critical feedback during the process and for having me in Rijnstate. Chedwa Pinto, thank you for coordinating the sampling in the Erasmus MC and for being involved in the process. Hetty Gerritse - Kattouw, thank you for scheduling all my appointments with Jan and making sure my account was extended so many times. Kim de Jong, my internal supervisor, thank you for helping out when things were difficult, for your feedback, and for your quick responses to my mails. Finally, I want to thank my parents, sister, friends and my housemates who helped me all through this process and provided me with the needed support.

Abstract

Hospitalized patients suffering from both somatic and psychiatric complaints have a worse prognosis than patients without psychiatric complaints. Therefore, specialized care in the form of Medical Psychiatric Units (MPU) is suggested. The present research was conducted to develop and validate an instrument which evaluates the level of disruptive behavior impeding somatic treatment, a focus point of an MPU. In Study 1, a first draft of the Inpatient Disruptive Behavior Index was designed; the InDiBI 1.0, focusing on severity of behavioral problems. In Study 2, a qualitative pilot study was conducted among nurses and physicians ($N = 14$) to evaluate content validity. Subsequently, the InDiBI 1.0 was adjusted and InDiBI 2.0 and 3.0 were developed, both focusing on manageability of disruptive behavior instead of the severity, with the InDiBI 2.0 being a multidimensional instrument, while the InDIBI 3.0 was unidimensional. In Study 3, nurses and physicians ($N = 54$) scored the 3 instruments using standardized vignettes. Feasibility was proven by few missing data. Intraclass correlations revealed good inter-rater reliability (ICC ≥ 0.53) and there were no signs that ward or job function affected scoring. Variety in scores between vignettes and positive correlations between instruments illustrated good construct validity. Regression analyses revealed sufficient levels ($R^2 \geq 0.44$) of explained variance between vignettes of all InDiBIs. These results imply that all versions are sufficient valid, thought with slightly different content. InDiBI 3.0 is recommended to use in practice as it was most directly related to the desired construct, least time-consuming to fill out, and preferred by the majority of the stakeholders.

Keywords: instrument development, validity, reliability, disruptive behavior, medical psychiatric unit

Table of Contents

Introduction

In hospitals, about 25% to 40% of the inpatients suffer from psychiatric comorbidity dealing with both psychiatric and somatic complaints (Hansen et al., 2001; Kathol, Saravay, Lobo, & Ormel, 2006; Silverstone, 1996; Wancata, Benda, Windhaber, & Nowotny, 2001). This group has a worse prognosis than patients who do not have psychiatric complaints (Kathol, Saravay, Lobo, & Ormel, 2006). An explanation for this worse prognosis is that the psychiatric complaints can lead to various forms of disruptive behavior or can be considered as disruptive in itself. Aggression, absconding, self-harm and medication refusal for example are linked to various negative outcomes such as injuries, treatment disruption, extended length of stay and it hinders patient recovery (Carmel & Hunter, 1989; Clark, Kiyimba, Bowers, Jarrett, & Mcfarlane, 1999; Kasper, Hoge, Feucht-Haviar, Cortina, & Cohen, 1997; Nijman & a Campo, 2002). In a recent extensive systematic review, the relationship between medical-psychiatric comorbidity and healthcare utilization and costs was examined. Results showed that this comorbidity was related to increased length of stay, medical costs and more rehospitalizations (Jansen, van Schijndel, van Waarde, & van Busschbach, 2018). These findings contributed to the recommendation of specialized care for this vulnerable group of patients in the form of a Medical Psychiatric Unit (MPU) (Kathol, Saravay, Lobo, & Ormel, 2006). Such MPUs consist of specialized staff able to provide complex somatic and psychiatric care simultaneously (Kathol, 1994; Van Waarde, Richter, Müller, & Verwey, 2004). It is hoped that the MPU can contain disruptive behavior more effectively and in that way allow more somatic care. Furthermore, the establishment of an MPU will lower the risk of stress among staff at general wards, as they are not all trained to work with complex psychiatric comorbidity. Other patients at the general ward also benefit from the presence of the MPU, as they do not have to share a room with such complex patients, and caregivers can distribute their time more equally to all patients instead of having to spend more time on the patients with the complex psychiatric comorbidity.

So far, there is only limited published evidence about the effectiveness of MPUs (Kishi & Kathol, 1999; Leue et al., 2010). One of the problems with establishing such proof is the lack of a validated outcome measure for one of the main outcomes of the MPU, which is *the level of disruptive behavior impeding the somatic treatment*. In this paper, we report on the development and validation of an observational rating scale to measure the manageability of such disruptive behavior: the 'Inpatient Disruptive Behavior Index' (InDiBI).

Previous research on the effectiveness of MPUs compared the MPU to general wards on multiple variables, such as medical diagnoses, psychiatric diagnoses, illness acuity, length of stay, medical service use, and exposure to hospital psychiatric interventions (Kishi & Kathol, 1999; Leue et al., 2010). These outcome variables certainly represent the ultimate goals of health care, but their expression will be influenced by much more than just optimal care at the MPU. For instance, the possibility to refer a patient back to the original general hospital unit or to any other follow up treatment facility, will determine the length of stay. It is doubtful whether a short stay is a sign of optimal care: if a patient is redirected to a unit without optimal facilities for somatic care because of the given disruptive behavior, a short stay may even be a sign of insufficient treatment. For instance, Leue and colleagues (2010) found that the length of stay of complex patients was longer at the MPU than at the medical ward. One could assume that better treatment at the MPU will result in lower overall cost (not just the cost in the particular hospital, but all medical costs associated with the patient). Moreover, one could argue that one should use an ultimate outcome of health care to compare results, such as quality of life and survival. Indeed, combining overall (societal) costs, quality of life and survival in cost per Quality Adjusted Life Years (QALY) is the preferred estimation of the cost effectiveness of any medical intervention (Drummond, Sculpher, Torrance, O'Brien, & Stoddart, 2005). On the other hand, one could question whether improving the health status of the patient and reducing costs are the only determinants of success of an MPU. As stated above, the care for other patients at the ward, the burden for the personnel, or even adequate delivery of treatment by well-trained personnel of patients with complex psychiatric comorbidity may all be warranted outcomes of the MPU. In addition, it can be questioned whether it will be possible to measure the long term outcome of quality of life, survival and overall costs of an intervention with an intervention like the MPU which is usually short. It might well be that any effect of the MPU will diffuse in other effects of the complex and dynamic treatment combinations typical for these patients. It would therefore be helpful to develop a valid outcome measure for the most important aim and facilities an MPU has, which is *containment of disruptive behavior impeding the somatic treatment*. This outcome measure can only be measured with an observational instrument, so it will not only measure the behavior of the patient, but it will also reflect the opinion of the personnel providing care about that behavior. Hereby, the instrument will give insight into two important facets of the MPU; the disruptive behavior of the patient and the perception of that behavior by the personnel as well as their interaction.

For the development of such an instrument, the definition of disruptive patient behavior should be operationalized. However, as it is such a broad construct, it is impossible to create a definition that encompasses every type of disruptive behavior. In the setting of an MPU, disruptive behavior can be defined as behavior which impedes or interrupts the delivery of care of the patient, but also that of other patients (Bowers et al., 2005). Bowers and colleagues (2005), for example, studied disruptive and dangerous behavior in patients on acute psychiatric wards in three European centers and included aggression, absconding, substance misuse and medication adherence in their definition. Nevertheless, behavior such as self-harm, disinhibition, calling out and not following instructions, could also be considered as disruptive. The present study is based on the types of disruptive behavior as established by a group of experts who determined the inclusion criteria for MPUs in a 'concept map' investigation (Caarls, Van Schijndel, Van Wijngaarden, & Van Busschbach, 2018, submitted).

The aim of this study was to fill up the gap and develop and validate a brief and simple instrument which could be used as outcome measurement to evaluater the effectiveness of an MPU in terms of containment of disruptive behavior. For the instrument, it was crucial that it could be administered easily by nursing staff and doctors in a tightly scheduled hospital environment. The development and validation of this instrument were the subject of three studies. During this process, the view of what the instrument was supposed to measure shifted from a focus on purely disruptive behavior to the 'manageability of disruptive behavior'. This means that the outcome of the MPU no longer is measure pure in terms of benefits for the patients, but for the staff and the organization as well. In this bundled study, the steps that led to this alteration are reported in detail. In Study 1, we started with the generation and the development of the items. In Study 2, we conducted a qualitative pilot study on the developed instrument to establish content validity. In addition, we improved the instrument accordingly to the feedback which resulted in three final versions of the questionnaire. In Study 3, we conducted quantitative research and consequently provided various validity evidence for the three instruments. The detailed hypotheses can be found in the overview sections of the separate studies.

Research Question

The 'research objective' was to arrive at a valid instrument to measure disruptive behavior in the context of an MPU. The resulting research question of this investigation therefore was: *"What is the most valid instrument that could be used as outcome measurement to evaluate the effectiveness of an MPU in terms of containment of disruptive*

behavior?" A valid questionnaire is defined as a questionnaire with good psychometric qualities, good content validity and it should be a practical measure to use, all in the context of an MPU setting.

Study 1: First Design

Overview

Study 1 was designed to establish the first design of the instrument and generate the items. First, the types of disruptive behavior that had to be included were established by an group of experts working at MPUs in various hospitals in the Netherlands. Next, these types of disruptive behavior were developed into items. Afterwards, these items were reviewed and the final items were established.

Method

Procedure. MPU experts provided the foundation for the new instrument. They decided on six dimensions that were considered as inpatient disruptive behavior in hospitals. This was done immediately after the determination of the inclusion criteria for MPUs in a 'concept map' investigation (Caarls, Van Schijndel, Van Wijngaarden, & Van Busschbach, 2018, submitted). The end-result of the concept map was a five cluster solution: 1. Staff competencies and organizational pre-requisites; 2. Patient context; 3. Patient characteristics; 4. Medical needs and capabilities; and 5. Psychiatric symptoms and behavioral problems. The experts operationalized this fifth cluster using six dimensions: agitation/aggression; suicidal behavior or deliberate self-harm; disinhibition; absconding or wandering behavior; calling out or moaning or making other sounds; and compliance with clinician instructions. Consequently, these dimensions were used to formulate items to arrive at an instrument. This process was done by the author of this paper and several other researchers of the Erasmus MC, who are involved in the evaluation of the forthcoming MPU at the Erasmus MC: Maarten van Schijndel, Jan van Busschbach, Chedwa Pinto. The process involved multiple rounds of discussions.

Results

The aim of the instrument was to measure the disruptive behavior symptoms and the severity level of this behavior. A first attempt to formulate the items was done by one of the authors (EvO), resulting in the InDiBI 0.1 (Appendix A). These items were based on the dimensions formed in the concept map, notably 'disruptive behavior'; which is the key variable that determines admission to an MPU. For each dimension, one item was included to assure the brevity of the instrument. Moreover, items were made as short and clear as possible. The items covering the dimensions of agitation/aggression and suicidal behavior or

deliberate self-harm were inspired by the items used in the Health of the Nation Outcome Scales (HoNOS) (Wing et al., 1998) but were adjusted to fit the needs of the InDiBI better.

Consequently, this list was discussed with one of the other researchers (JvB) and improvements were made, leading to the InDiBI 0.2 (Appendix B). This instrument was based on the dimensions that were provided as well as on features of the validated HoNOS (Brooks, 2009; Mulder et al., 2004; Wing et al., 1998). However, unlike the HoNOS, the InDiBI 0.2 only included 4-point scales and 3-point scales instead of 5-point scale. This choice was made, as the 5-point scale seemed too detailed to be reliable for the interrater reliability, especially regarding the more severe answer options. For the dimensions 'disinhibition' and 'compliance with clinician instructions', a 3-point scale covered the various levels of 'disruptive behavior', while for the other dimensions a 4-point scale was needed. Subjective wording in the answer levels such as 'sometimes, most of the time, every now and then' were avoided if possible.

During a successive discussion round with another researcher (CP), one additional dimension was mentioned; 'asking unnecessarily for attention'. This type of behavior did not seem to be covered by the other dimensions yet and was therefore added as an item, resulting in the InDiBI 0.3 (Appendix C). This third version was considered complete and was then used in the qualitative pilot study in Study 2.

Study 2: Qualitative Evaluation
Overview
Study 2 was designed to establish content validity of the instrument by asking participants whether they agreed that the items covered the content of 'disruptive behavior'. Their feedback was written down in a qualitative report. Content validity was hypothesized to be met if the participants agreed on the items. According to the results, the instrument was adjusted and two extra versions were developed. This resulted into three different versions of the InDiBI.

Method
Participants. The sample ($N = 14$, see Table 1) included nursing staff and physicians who worked at three different units at Erasmus MC in Rotterdam, the Netherlands; Erasmus MC Cancer Institute – Location Daniel den Hoed, Department of Hematology, Erasmus MC Unit P3 – Psychiatry and Somatic Comorbidity and Pregnancy-Related Psychiatry, and the Department of Internal Medicine. Participants were selected to be included based on a combination of stratified and convenience sampling. It was a stratified sample, because the

sample was forced to include participants of predetermined subgroups of the target population; nursing staff and physicians at three different unites. It was convenience sampling, because participants were selected based on their availability and willingness during the assessment time. All participants were informed about the purpose of the research and asked to sign an informed consent form. Participation was without incentive. Under Dutch law no medical ethical approval was necessary, as no intervention took place, nor could the interview considered to be laborious.

Table 1

Sample characteristics

Variable	*n*	*%*
Gender		
Male	2	14.29
Female	12	85.71
Job function		
Nurse	5	35.71
Clinician	8	57.14
Unknown[a]	1	7.14

a. This participant did not fill in the job function

Research design. The first version of the InDiBI was used in a pilot study in clinical practice, in which 14 participants provided feedback on the InDiBI. The design can be best classified as a 'cross-sectional study design' since all participants followed the same procedure and data was collected at a single point in time.

Measures.

Inpatient Disruptive Behavior Index 0.3 (InDiBI 0.3).

The InDiBI 0.3 is an observational rating scale that consists of seven dimensions of disruptive behavior of patients in hospital settings. The questionnaire is designed for caregivers working in hospitals such as nursing staff and doctors and is made in the Dutch language. For the answer options, each construct has their own specific scale. For all constructs, answer option 1 means that the type of disruptive behavior is not present. Furthermore, for all items, the rating scale is ascending in severity, meaning that answer option 3 indicates more disruptive behavior than answer option 2 and so on. The difference among the constructs is the amount of answer options. For some constructs, a 3-point scale was sufficient, while for other constructs a 4-point scale was needed.

Procedure. In the qualitative pilot study, an examiner (EVO) interviewed participants individually at their workplace. An interview lasted 10 minutes on average. The interviews were semi-structured and the items of the questionnaire were discussed one by one. This way,

the feedback was already clearly ordered. Participants were asked to fill out the instrument based on a patient exhibiting disruptive behavior from their current or past caseload and to give feedback on the clinical utility of the instrument. They were asked to comment on wording, clarity, and completeness of the preliminary InDiBI. The respondents who filled out the forms (Appendix D) by themselves in their own time also delivered clear output per question. Therefore, an extensive coding scheme was regarded as redundant.

The analysis was done according to the following steps. First, the feedback of all participants was noted down per item and summarized. Consequently, this feedback was used to improve the items. This phase was included to examine the face validity, content validity, and the feasibility the InDiBI. The new improved items were written down under the original pilot items. Next, all the additional comments were included. These have been split up into two sections; general comments and missing categories. To ensure that no comments were forgotten, it was checked per participant whether all their feedback was incorporated.

Results

The pilot study resulted into more insight in how to develop the questionnaire to suit the ultimate goal; measuring the effectiveness of an MPU in terms of the burden of disruptive behavior on the personnel. The aim was to establish content validity of the questionnaire, which was evaluated by analyzing all the feedback of the participants. An extensive report of the results and discussion of the qualitative analysis can be found in appendix E.

It turned out that the items as proposed were reviewed as unclear. Almost all participants had comments about the answer options. They were seen as incomplete, unclear and sometimes even contradicting. Each item either had a 3- or 4-point scale ascending in severity of the type of disruptive behavior, but this was not always recognized. Participants also struggled in choosing the answer option that fitted the behavior of their patient best. Furthermore, many specific comments were given directed to the individual items or answer options. To improve content validity, these have been taken into account and a new version of the InDiBI 1.0 (Appendix F) was developed.

After the improvement of the InDiBI 1.0 however, the questionnaire still did not seem to incorporate all the feedback that was given by the participants. Some also opted for extra categories, including delirious and catatonic behavior. These categories indeed were not yet covered in the index, but do certainly represent disruptive behavior. Other categories that were initiated such as cognition, intoxication, the patients' state, and severity of illness were not added, as these categories do not always have to lead to disruptive behavior. Verbal

aggression was also mentioned to be missing, but this type of behavior can be scored in the item about agitation and aggression.

In addition, the qualitative pilot study and more discussion rounds questioned whether the focus on the severity of disruptive behavior would be the correct way to measure the effectiveness of the MPU. The thought was that this approach could reveal that these symptoms would decrease more or more rapidly at an MPU compared to a general ward. This would imply that the MPU would be effective in the containment of disruptive behavior. However, it might very well be that there will not be a difference between the severity of disruptive behavior among a general ward and an MPU considering the short length of stay in hospitals. Thus, the MPU is not only about decreasing disruptive behavior and improving the health status of the patient; it is about more than that. The severity of disruptive behavior on its own does not determine directly how effective the MPU is in the treatment of patients. Less disruptive behavior means less treatment interference. However, the success of treatment also depends on the quality of care provided by the care givers. This insight led to the realization to focus on the manageability of disruptive behavior. If a nurse or physician feels like he or she is able to manage the disruptive behavior, he or she is able to provide high quality care and more attention for the somatic care leading to more effective treatment, which is the ultimate goal of the MPU. Consequently, the success of an MPU is ultimately in the hands of the care givers. It could even be argued that the MPU is not only created for the patient, but for the nurses and physicians who have to deal with the patients. The MPU benefits care givers at general wards by taking over patients with psychiatric comorbidity which will result in less burden for the general ward. Furthermore, the other patients at the general ward benefit as well, because they can receive more attention from the care givers and are less disturbed by the patients with psychiatric comorbidity. Based on this new perspective, the InDiBI 2.0 was developed, in which the disruptive behavior categories could be scored in terms of manageability (Appendix G).

Though, after the realization of this second version, other questions were raised. Why would the disruptive behavior have to be categorized? What would it mean if a caregiver would score a patient as not manageable on two categories, but as manageable on the other seven categories? Is this patient manageable or not? In other words, it would be difficult to come up with a cut off score to decide whether the disruptive behaviour would be seen as manageable and when it would be seen as not manageable. Therefore, a third version of the InDiBI (Appendix H) was developed; the InDiBI 3.0, with only one question: 'Does the patient exhibit disruptive behavior and is this patient manageable?' and a sub question in

which the care giver could score what behavior was seen as disruptive. In this third version, the problem of a subjective cut off score was solved; the patient could be either scored as being manageable or not. Furthermore, in the InDiBI 3.0, there is more space for patient characteristics and behavior that make a patient seen as manageable or not, which cannot be grasped in a distinct category.

Consequently, the three versions of the InDiBI could be used in the final step of this research; Study 3, which focused on the validation of the instruments and examined which version is preferred by the participants who will have to use the instrument.

Study 3: Instrument Validation

Overview

Study 3 was included to validate the three questionnaires that were developed in the previous studies. This was achieved by doing quantitative research where a sample of nursing staff and physicians filled out a research bundle. They scored five standardized vignettes and one of their own patients using the three developed instruments. Consequently, the obtained data was used for analysis to evaluate hypotheses regarding feasibility, reliability, and validity of the questionnaires.

Feasibility included data completeness on the items of the questionnaires and the preferred version by the participants. As indication of a concrete questionnaire, it was hypothesized that there would be few missing data and that they would be missing at random. Furthermore, it was hypothesized that the InDiBI 3.0 would be preferred by the majority, as it was the version made to be most convenient to fill out.

Reliability analyses were performed to evaluate consistency of the data reported. As the InDiBI is an observational instrument, it should have a reasonable 'inter-rater reliability' as well. An indication of reasonable inter-rater reliability would be expressed by good to excellent intraclass correlations between scores on the patient vignettes rated by different participants. Note that the inter-rater reliability is difficult to assess using the scores of the 'real life patients' as one is then evaluate different patients or the same patients at different moments. Furthermore, it was checked whether the ward where the 'rater' is working and the job function of the rater had an effect on scoring.

Validity analyses were included to evaluate whether the questionnaire measured what they were intended to measure. As an indication of 'construct validity', moderate or higher correlations of the mean scores of the three versions of the InDiBI were expected. Furthermore, validity of the three versions of the InDiBI was checked by comparing mean

scores in a mixed model analysis. Construct validity was tested using the hypothesis that the three versions would be able to significantly distinguish the five vignettes. Lastly, it was checked what the proportion of explained variance was of each questionnaire. This was done to evaluate whether the instruments were useful in differentiating various patients.

Method

       Participants. Participants were recruited in two hospitals in the Netherlands; Erasmus MC (Rotterdam) and Rijnstate (Arnhem). The same combination of stratified and opportunity sampling as used in the development phase was used. In the Erasmus MC, three psychiatric wards and two general wards, internal oncology and internal medicine (*interne ouderengeneeskunde*), were included. In the Rijnstate hospital, the medical psychiatric unit (MPU) was included. Inclusion criteria for participants were working at one of those wards as nursing staff, psychiatrist, physician or resident (*arts in opleiding tot specialist - AIOS*).

       The aim was to obtain at least 10 participants (five nurses and five physicians) per type of hospital ward (psychiatric ward, somatic ward or MPU). Participants were given a letter with all information about the research (Appendix I) Furthermore, they were asked to fill in an informed consent form (Appendix J) and a form in which they had to fill in some basic demographical information (Appendix K).

       In Table 2, the sample characteristics are displayed ($N = 54$). It gives a breakdown of the participants by ward and profession, as well as gender, age and years of experience. The average age of the participants was 39.61 ($SD = 12.09$). The average years of experience was 11.25 ($SD = 11.27$).

Table 2

Sample characteristics

Variable	M	SD
Age[a]	39.61	12.09
Years of experience	11.25	11.27

Variable	n	%
Gender		
Male	22	40.7
Female	31	57.4
Missing	1	1.9
Hospital		
Erasmus MC	34	63.0
Rijnstate	20	37.0
Ward		
Psychiatry	31	57.4
Internal oncology	5	9.3
Internal medicine	1	1.9
MPU	17	31.5
Job function		
Nurse	10	18.5
Nurse in training	1	1.9
Psychiatric nurse	15	27.8
Psychiatrist	8	14.8
Internist	1	1.9
AIOS EMC psychiatry	12	22.2
AIOS EMC somatic	4	7.4
AIOS RIJN psychiatry	3	5.6

a. Age: $n = 51$

Research design. The design could be best classified as a 'cross-sectional study design' since all participants followed the same procedure and data was collected at a single point in time (May – August 2018). Participants filled out all three versions of the InDiBI, using five vignettes which were written out on sheets (Appendix L). Using general linear models we want to explore whether the three versions of the InDiBI give different results and whether participants with different background give different results.

Measures.

Standardized vignettes.

The vignettes were developed with the aim to portray realistic patients as seen in the hospital and the casus were evaluated by two psychiatrists working in the hospital (MVS and CP). The five vignettes were highly heterogeneous to investigate whether the InDiBI could be applied to various patient situations. The vignettes were not only meant to include various types of patients regarding somatic and psychiatric complaints, but foremost to display variation in severity of disruptive behavior that was expressed.

General information participant form.

To obtain a basic impression of the participants, date of birth and gender were asked. Furthermore, it was registered in what hospital they worked and at which unit as well as their job function and years of experience in this job function. The form also included one question in which the participant could indicate which version of the InDiBI he or she preferred to use.

Inpatient Disruptive Behavior Index 1.0 (InDiBI 1.0).

The InDiBI 1.0 is the instrument that was developed in the first phase of this study. It is an observational rating scale that measures seven dimensions of disruptive behavior of patients in hospital settings, using a mixture of frequency and severity as response levels. The questionnaire was designed for caregivers working in hospitals such as nursing staff and doctors and was made in the Dutch language.

Inpatient Disruptive Behavior Index 2.0 (InDiBI 2.0).

The InDiBI 2.0 is a more elaborate observational instrument which measures the manageability of disruptive behavior of patients for the personnel. The instrument includes nine dimensions of disruptive behavior. The response levels are unidimensional and express the level of manageability of the disruptive behavior. The questionnaire was designed for caregivers in the hospital such as nursing staff and doctors and was made in the Dutch language.

Inpatient Disruptive Behavior Index 3.0 (InDiBI 3.0).

The InDiBI 3.0 is a simplified observational instrument which measures the overall assessment of the manageability of disruptive behavior of patients for the personnel. The instrument includes one general question about the manageability of the patient with three levels to score. In the second question, various types of disruptive behavior can be scored. The questionnaire was designed for caregivers in the hospital such as nursing staff and doctors and was made in the Dutch language.

Procedure. In the validation phase of the InDiBI, nursing staff and physicians of two different hospitals in the Netherlands were asked to score both five patient vignettes and one real life patient using the three versions of the InDiBI. They did this by filling in the research bundle, which was handed to them in print.

Data analysis. SPSS statistical software version 24.0 (IBM Corp, 2016) was used for all analyses. The level of statistical significance was set at a *p*-value of 0.05. Missing data were handled through mean imputation if more than half of the items per questionnaire per case (vignette or real patient) was filled out. As the three questionnaires had different outcome

scales, scores were transformed to a scale from 0 to 3, to allow comparisons between the three versions. Furthermore, participants who indicated to work at the internal oncology and internal medicine wards were merged together as working at a somatic ward, in order to compare this group to the participants working at a psychiatric ward and the participants who work at the MPU. For feasibility, a missing value analysis was performed to evaluate data completeness. In addition, the preferred InDiBI version was investigated by creating a frequencies table.

Intraclass correlation (ICC) analyses were used for inter-rater reliability, as this analysis reflects both the agreement and correlation of multiple measures rated by multiple raters. The data was first transposed, so that the rows would display the vignettes and the columns would display the participants. The appropriate ICC estimation was selected by following the guidelines set by Koo and Li (2016) and Shrout and Fleiss (1979). In this research, it contained a two-way random effects, consistency, single rater model. It was a two-way model as both the vignettes and the raters were a sample of all possible vignettes and raters, and thus raters and vignettes were considered representative from a larger population and the ICC were meant to generalize to that population. In other words, it was based on the idea of consistency over raters, as it was important that raters provide scores that are similar in at least the rank order. It was considered a 'single rater case', as the aim was to generalize the reliability to a single rater in future. This is the case, as in practice only one or a few nurses or doctors will score a patient which will lead to a decision. (Hallgren, 2012). The interpretation of the ICC values according to commonly-cited Cicchetti (1994) is as following; values less than 0.40 indicate poor inter-rater reliability, values between 0.40 and 0.59 indicate fair reliability, values between 0.60 and 0.74 indicate good reliability and values between 0.75 and 1.0 indicate excellent reliability.

In order to test if wards and job function had an effect on the scoring, mixed model analyses were used, treating job and ward as fixed effects. To facilitate a visual inspection of the validity, mean scores of all vignettes per version of the InDiBI were retrieved by the creation of descriptive tables and a figure with standardized scores. Next, bivariate correlation analyses were run to evaluate whether the three InDiBI versions correlated. The interpretation of these values is based on the criteria set by Hinkle, Wiersma, and Jurs (2003). Values between 0.00 and 0.30 are negligible, values between 0.30 and 0.50 are low, values between 0.50 and 0.70 are moderate, values between 0.70 and 0.90 are high, and values between 0.90 and 1.00 are very high. To check whether the versions would be able to significantly

distinguish the vignettes, a mixed model analysis was used. Lastly, a regression analysis was run to find out whether the InDiBIs predicted the scores of the vignettes well.

Results

      Assumptions. For the regression analyses, several assumptions had to be checked. Normality of the outcome variables; the mean scores per questionnaire per vignette, was assessed using normal P-P plots and the data turned out to be normal. Homoscedasticity of the residuals was checked with scatterplots and was considered to be met. Furthermore, in all three regression analyses, tolerance was > 0.1 and the variance inflation factor < 10, thus the assumption of multicollinearity was met as well.

      Data completeness.

      Missing values were checked for all questions in the questionnaires for all vignettes and the real patient. In Table 3 can be seen that 64.8% had no missing values and 20.4% had only one missing value. Two participants had 31 missing values, because they left the items for the real patient blank. The missing values were also checked per item. There were five items that had three missing values. All other items had fewer missing values.

Table 3

Missing values on participant level

Values	*n*	%
0	35	64.8
1	11	20.4
2	1	1.9
3	2	3.7
4	1	1.9
5	1	1.9
17	1	1.9
31	2	3.7
Total	54	100.0

      Preferred InDiBI version.

      Table 4 shows that 51 participants answered the question about the favorite InDiBI and 3 participants left this question blank. Of the participants who did fill out the question, 49.0% preferred the InDiBI 3.0, 29.4% preferred InDiBI 2.0 and 21.6% preferred the InDiBI 1.0.

Table 4

Favorite InDiBI

Variable	*n*	%	Valid %
InDiBI 1	11	20.4	21.6
InDiBI 2	15	27.8	29.4
InDiBI 3	25	46.3	49.0
Total	51	94.4	100.0
Missing system	3	5.6	21.6
Total	54	100.0	

Reliability.

Inter-rater reliability.

The results of the intraclass correlation coefficients of each version of the InDiBI over raters can be seen in Table 5. For the InDiBI 1.0, one vignette was excluded as there was a missing value for that vignette among one of the raters. Thus, the ICC was computed with 54 raters and four ratees (the vignettes) for the InDiBI 1.0 and with five ratees (vignettes) for the InDiBI 2.0 and 3.0. The ICCs of interest were respectively .78, .68, and .53. The average ratings were remarkably higher; respectively 1.00, 0.99, and 0.98. Note that these 'group' outcomes might be less relevant here. See also the discussion about these high value in the discussion section.

Table 5

Reliability statistics

Variable	ICC (95% CI)[a]
InDiBI 1.0 Single Measures[a]	.78 (.52-.98)
InDiBI 1.0 Average Measures[a]	1.00 (.98-1.00)
InDiBI 2.0 Single Measures[a]	.68 (.42-.95)
InDiBI 2.0 Average Measures[a]	.99 (.98-.1.00)
InDiBI 3.0 Single Measures [a]	.53 (.28-.90)
InDiBI 3.0 Average Measures[a]	.98 (.95-.1.00)

Note. Two-way random effects model where both people effects and measures effects are assumed random.
a. $p < .00$ for all values

Variation in scoring between wards and job function.

Table 6 and 7 display how many vignettes per ward and per job function were scored. Consequently, a mixed linear model was run to estimate variation in scoring between wards and job function. All *F*-values for ward and job function on any of the three questionnaires turned out to be insignificant, with *p*-values being .23 and higher.

Table 6

Vignettes rated per type of ward

Ward	n	%	Valid %	Cumulative %
Psychiatry	155	57.4	57.4	57.4
Somatic	30	11.1	11.1	68.5
MPU	85	31.5	31.5	100.0
Total	270	100.0	100.0	

Table 7

Vignettes rated per type of job function

Job function	n	%	Valid %	Cumulative %
Nurse	50	18.5	18.5	18.5
Psychiatric nurse	75	27.8	27.8	46.3
Internist	5	1.9	1.9	48.1
Psychiatrist	40	14.8	14.8	63.0
Doctor assistant EMC psychiatry	60	22.2	22.2	85.2
Nurse in training	5	1.9	1.9	87.0
Doctor assistant EMC somatic	20	7.4	7.4	94.4
Doctor assistant RIJN psychiatry	15	5.6	5.6	100.0
Total	270	100.0	100.0	

Validity. Tables 8, 9, and 10 show the mean scores on the three versions of the InDiBI per vignette and the patient participants had encountered in their own ward. The score range is between 0 and 1 with a higher score meaning a higher amount of *disruptive behavior* in InDiBI 1.0. For the InDiBI 2.0 and 3.0, a higher score indicates that the patient is evaluated as more difficult to *manage*. On each questionnaire, the obtained mean scores vary, indicating that differences in level of severity among the vignettes and real patient are present. On all three InDiBI versions vignette 'De Jonker' is rated as most disruptive and vignette 'Veen' as least disruptive. The values for the InDiBI 1.0 and 2.0 are almost similar. Furthermore, InDiBI 3.0 follows the same trend in scores as the other two InDiBIs, but with a constant difference as the scale is not made fully compatible. The standard deviation was highest for the real patient in all three questionnaires, which makes sense as the participants all scored their own unique patient.

Table 8

Mean scores InDiBI 1.0

Case	n	Min	Max	M	SD
Hassan	53	.00	.64	.44	.13
De Jonker	54	.12	.74	.53	.12
Veen	54	.00	.38	.18	.08
Sardjoe	54	.05	.71	.28	.16
Steenbergen	54	.29	.86	.49	.10
Real patient	52	.00	.81	.34	.20

Table 9

Mean scores InDiBI 2.0

Case	n	Min	Max	M	SD
Hassan	54	.06	.65	.30	.13
De Jonker	54	.33	.89	.58	.14
Veen	54	.00	.33	.12	.07
Sardjoe	54	.13	.78	.34	.15
Steenbergen	54	.17	.85	.44	.15
Real patient	52	.00	.85	.31	.20

Table 10

Mean scores InDiBI 3.0

Case	n	Min	Max	M	SD
Hassan	54	.00	1.00	.54	.20
De Jonker	54	.50	1.00	.91	.20
Veen	54	.00	1.00	.44	.19
Sardjoe	54	.50	1.00	.87	.22
Steenbergen	54	.50	1.00	.80	.25
Real patient	52	.00	1.00	.54	.31

In Figure 1 below, the standardized mean scores of the five different vignettes are displayed for each questionnaire separately. By using z-scores, there has been controlled for the diverse scales used in the three questionnaires.
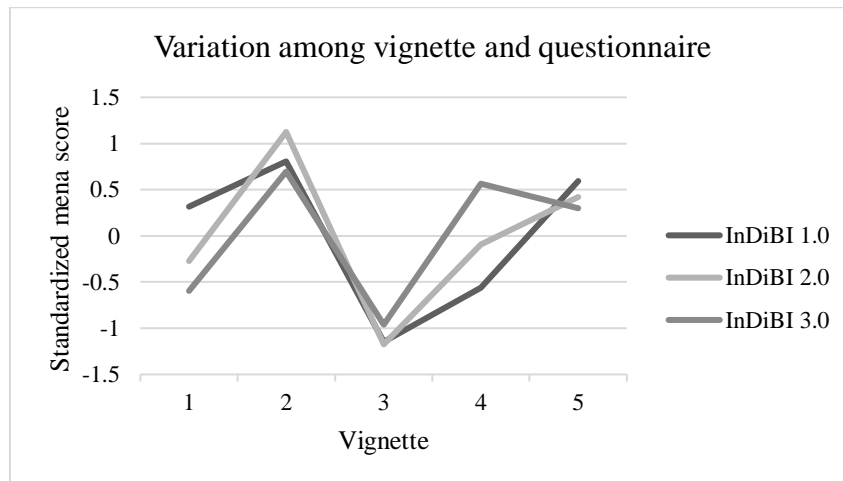

Figure 1. Variation among vignette and questionnaire

In addition, a correlation analysis of the three InDiBIs was done. In this analysis, all vignettes and the real patient were included separately. The sample size differed as some people did not fil out a certain question in one or multiple vignettes and therefore these submissions were left out. For the correlation between InDiBI 1.0 and InDiBI 2.0, $N = 269$, for the other correlations, $N = 270$. All correlations were found to be positive. The correlation between InDiBI 1.0 and 2.0 was $r = .76$, $p = < .00$. The correlation between InDiBI 1.0 and

3.0 was $r = .36$, $p = < .001$. The correlation between InDiBI 2.0 and 3.0 was $r = .68$, $p = < .00$. Note that the correlation between InDiBI 3.0 with the other questionnaires is most likely affected by the small scale of the InDiBI 3.0, making it almost a point-biserial correlation.

The next step was to check whether the five vignettes were rated differently. This should be the case as they were designed to display various levels of disruptive behavior. For this analysis, the sample of rated vignettes was $N = 809$. As can be seen in the outcome of the mixed model analysis, both vignette $F(4, 755.01) = 165.86$, $p = .00$ and version $F(2, 755.01) = 415.32$, $p = .00$ effects were significant. Thus, the vignette characteristics and the different versions of the InDiBI are statistically significant predictors of the scores given. In Table 11, all vignettes have a significant ($p = .00$ for all) higher or lower mean score compared to vignette five, which was considered the reference vignette (Steenbergen). Furthermore, the mean scores on the InDiBI 1.0 and 2.0 were significantly ($p = .00$; $p = .00$) lower compared to the mean scores on the InDiBI 3.0. The Wald Z statistic is significant, but the estimated variance of the intercept is .00.

Table 11

Estimates of fixed effects and covariance parameters[a]

Parameter	Estimate	SE	df	t	p	95% CI	
						Lower bound	Upper bound
Intercept	.80	.02	329.32	46.75	.00	.77	.84
Vignette Hassan	-.15	.02	755.14	-8.31	.00	-.18	-.11
Vignette De Jonker	.10	.02	755.04	5.50	.00	.06	.13
Vignette Veen	-.33	.02	755.04	-18.61	.00	-.36	-.29
Vignette Sardjoe	-.08	.02	755.04	-4.49	.00	-.11	-.04
Vignette Steenbergen	0[b]	0	.	.	.	.	.
InDiBI 1.0	-.33	.01	755.10	-23.89	.00	-.35	-.30
InDiBI 2.0	-.35	.01	755.04	-25.90	.00	-.38	-.327
InDiBI 3.0	0[b]	0	.	.	.	.	.

Parameter	Estimate	SE	Wald Z.	p	95% CI	
					Lower bound	Upper bound
Residual	.03	.00	19.43	.00	.02	.03
Intercept [subject = ID] Variance	.00	.00	3.70	.00	.00	.01

a. Dependent Variable: InDiBI score
b. This parameter is set to zero because it is redundant.

Explained variance per questionnaire.

To evaluate the variance explained by the five case patients per questionnaire, multiple linear regression analyses were performed. The real patient was excluded from this analysis as this one was unique for all participants. As can be seen in Table 12, the R^2 for InDiBI 1.0, 2.0 and 3.0 were subsequently 0.55, 0.58 and 0.44. That means that a substantial part of the

variance can be attributed by differences between vignettes, and thus the questionnaires are able to present this variance in a reasonable amount besides the measurement error.

Table 12

Model Summary

Version	R	R^2	Adjusted R^2	SE of the Estimate
InDiBI 1.0	.74[a]	.55	.54	.12
InDiBI 2.0	.76[a]	.58	.58	.13
InDiBI 3.0	.66[a]	.44	.43	.21

a. Predictors: (Constant), vignette2, vignette3, vignette4, vignette5

Discussion

In this threefold study, the development and validation of an instrument that could be used as outcome measurement to evaluate the effectiveness of an MPU in the containment of disruptive behavior was evaluated. In Study 1 and 2, three questionnaires were created, and their content validity was assessed in study 2. In Study 3, hypotheses on the feasibility, reliability, and validity of the final three versions of the InDiBI were evaluated, to investigate which version of the instrument could best be used in practice and further research.

The hypothesis on content validity was not met, as participants reported many ambiguities, inconsistencies and points for improvement. Therefore, more effort was taken to improve the content of the questionnaire, resulting in version 1 of the InDiBI. The feedback also shed light on another approach to measure the effectiveness of an MPU, questioning whether the disruptive behavior was seen as manageable by the care givers. To measure this main objective of an MPU, InDiBI 2.0 and 3.0 were developed.

The hypothesis to establish feasibility was met as the data contained few missing values. Furthermore, no patterns were found in the missing data. Thus, it can be assumed that the missing values were at random. These results indicated that the three versions of the questionnaires were well understood and consequently all versions seem to be concrete instruments. The other hypothesis, which assumed that InDiBI 3.0 would be the preferred version, was met as well, as this version was preferred by almost half of the sample.

Reliability was checked using various analyses. For an indication of reasonable inter-rater reliability, good to excellent ICC values were desired. InDiBI 1.0 and 2.0 met this hypothesis, meaning that patients were rated similarly across participants. InDiBI 3.0, however, only reached fair inter-rater reliability (Cicchetti, 1994). Contrastingly, the average measures ICC values for all versions reached excellent inter-rater reliability. Nonetheless,

these values are of less meaning, as they show the reliability for the generalization of the results to another group of raters instead of single raters, which will not happen in practice (Shrout & Fleiss, 1994). In addition, for all three InDiBI versions, no significant differences were found in scoring between wards or job functions. This suggests that people working at different wards or people with different job functions score the questionnaires in the same way, which was wished for in order to develop a reliable measurement instrument.

Construct validity was assessed in multiple ways. Results revealed variation among the different vignettes, indicating that the questionnaires were able to distinguish these well. The tables of mean scores showed that InDiBI 3.0 followed the same pattern as the other two InDiBIs when scores were standardized, only with higher scores. The difference can be explained as the InDiBI 3.0 only consisted of one question, resulting in a statistical score of either 0.0, 0.5 or 1.0. Thus, if a participant thinks that a patient is showing disruptive behavior, this immediately results in a value of 0.5 or 1.0. Therefore, the scores were more extreme than the scores obtained in the InDiBI 1.0 and 2.0, since scores of these questionnaires consist of multiple components. To overcome this problem, Figure 1 was made with standardized scores which showed that the values were indeed more similar. In line with the observed results, the correlation analysis showed positive and significant correlations. The hypothesis to find moderate or higher correlations of the mean scores of the three versions of the InDiBI was not entirely met, as the correlation between the InDiBI 1.0 and 3.0 was low. However, correlations between version 1.0 and 3.0 and between version 2.0 and 3.0 were considered high and moderate. These results showed that severity of disruptive behavior is indeed linked to the manageability of this disruptive behavior. Furthermore, the results reflect that both InDiBI 2.0 and 3.0 measured manageability and not the level disruptive behavior as in version 1.0. Although the hypothesis was not entirely met, this can be explained as version 1.0 measured severity with multiple questions and version 3.0 measured manageability with only one question. That means that much variance is lost and thus less variance can be explained. Next, the hypothesis that the three versions would be able to significantly distinguish the five vignettes was met, as the mean scores were significantly different for both vignette and version. This indicates that the InDiBIs were able to differentiate between various levels of disruptive behavior or manageability of disruptive behavior. Secondly, significant different mean scores were found between InDiBI 1.0 and InDiBI 2.0 compared to InDiBI 3.0. This can be explained by the constant trend of lower mean scores on the InDiBI 1.0 and InDiBI 2.0 compared the InDiBI 3.0 due to the different scales that were used.

Regression analyses revealed that the models significantly predicted the InDiBI scores. A substantial part of the variance can be attributed by differences between vignettes. Version 1.0 and 2.0 differentiate better than InDiBI 3.0, but this was expected the InDiBI 3.0 consists of only one question with three answer options, which reduces the possibility of variance considerable. This is not seen as problematic, as the aim was to measure the containment of disruptive behavior on group level in a research context and not on individual level in a clinical context. The use of only one question is sufficient and more efficient to measure on group level; more questions would be simply redundant.

A number of potential limitations need to be considered. In Study 2, a limitation was that most participants filled out the feedback questionnaire themselves in their own time. This might have led to misunderstanding of the aim or use of the questionnaire. It became clear from multiple respondents that they did not understand for what purpose the InDiBI would be used. They mentioned for example that it was unclear for them whether they had to score the behavior as being disruptive for other patients or for the personnel or for the patients themselves. Consequently, the feedback was sometimes slightly limited, misplaced or unhelpful, thus the pilot phase has not reached optimal results for all respondents. Still, there was enough useful feedback, so this was not seen as a major threat.

Moreover, after the pilot study, the three versions of the InDiBI were made to be used in the validation study. These newly developed versions have not been reviewed again by experts. However, the improvements were based on their feedback, thus this second evaluation loop was not seen as a necessary step. Furthermore, it is unknown how many participants were recruited as this was done by multiple researchers and we lost track of the participants. The only underrepresented group in the sample are people working at the somatic ward. Therefore, the results comparing this ward should be interpreted with caution.

Next, in most validation studies, existing questionnaires often have been used to establish convergent validity. This was not possible in this research, as other questionnaires did not seem to overlap the new measurements. To overcome this problem, we compared the three versions of the instrument (InDiBI 1.0, 2.0, and 3.0).

Another shortcoming was the use of vignettes on paper as the scoring might have been more based on reading comprehension and interpreting text instead of on their true feelings and experiences with patients. Consequently, the results might be biased by this possible indirect effect. Other studies (Jones, Gerrity, & Earp, 1990; Rudwaleit et al., 2009) have also been using paper patients and concluded that the clinical information might not be detailed enough to make profound conclusions. Furthermore, the assessment of paper patients

compared to real patients might lead to different conclusions. Video vignettes might have overcome some of these difficulties, but this was not possible within the scope of this research. The use of standardized paper vignettes was most feasible and enabled the research to include 54 participants. Evidently, it would have been almost impossible to find 54 nurses and doctors in the hospital who would rate the same real patient.

Lastly, in the analyses became clear that the ward does not have any significant influence on scoring, which implies that people working at the MPU do not have different evaluations of disruptive behavior, while it was expected that they would attribute lower scores to the vignettes and real patients. It remains unclear how this result was obtained, but there are three possible explanations. The first one is that the written vignettes on paper were not realistic enough and hereby, participants were pointed towards certain scores. The second possibility is that the instruments are not specific enough to find differences between participants. The third explanation is that there is indeed no difference between people working at different wards, meaning that people working at the MPU are not more effective in managing disruptive behavior. More research in a hospital where nurses and physicians score the same real patients over time should give more insight into this ambiguous result.

A strength of this research was the elaborate focus on the design of the instrument. A lot of thought was given into this process to ensure that the instrument would be fitting to the needs; creating an outcome measurement to evaluate the effectiveness of an MPU. Furthermore, the continuous discussion rounds, feedback and critical thinking was a major strength in this research. Only through these circumstances, there was room for shifting the focus of the instrument from severity to that of disruptive behavior to manageability of behavior, which led to three separate versions of the InDiBI. Consequently, this resulted in a more profound comparison analysis. Moreover, this study involved both qualitative and quantitative analysis, which contributed to the development and validation of the questionnaires. Another strength was the combination of using both vignettes and a real patient in the validation study. The vignettes were useful for comparing between participants and analyzing reliability and validity, while the scoring of the real patient showed the InDiBIs actually could be applied to patients in the hospital, which was the ultimate aim.

In conclusion, concerning feasibility and reliability, all questionnaires had good results. Regarding validity, it became clear that severity of disruptive behavior was linked to manageability of this behavior. The results indicated that a higher level of disruptive behavior also means that the patient is more difficult to manage. InDiBI 1.0 and 2.0 had the highest explained variance, but also consisted of more questions. As the loss of explained variance is

not much, the InDiBI 3.0 can be regarded as a good questionnaire to measure the manageability of disruptive behavior of patients in hospital at group level. The InDiBI 1.0 and 2.0 consist of subsequently seven and nine questions, while the InDiBI 3.0 only consists of one question and a sub question. Therefore, it is more practical to fill out, while it saves time, and its interpretation is easier as the single item is directly linked to the construct of interest; 'containment of disruptive behavior'. This behavior is not limited to several categories as in the InDiBI 1.0 and 2.0, so this ensures that no type of disruptive behavior will be missed out. Furthermore, high scores on the InDiBI 1.0 and 2.0 the several categories might imply that the disruptive behavior is severe or unmanageably, but it would remain arbitrary how to weigh the categories and scores and what the cut-off score would be. This is not a problem in the InDiBI 3.0, cause in this questionnaire it is directly asked whether the observer thinks that the patient is manageable or not. Conclusively, all InDiBIs turned out to be reliable and valid questionnaires. The InDiBI 3.0 is however recommended to use in practice. This version with only one question can replace the other two versions which consist of a whole list of items. Most importantly, this version comes most close to measuring the ultimate goal of evaluating the effectiveness of the MPU; the containment of disruptive behavior. It is short and easy to use, which is especially important in a tightly scheduled hospital setting. Moreover, the majority of participants indicated this version as their favorite.

The findings of this study are beneficial for other research as well. Firstly, this study showed the complexity to find out how to exactly measure the concept of the effectiveness of an MPU. In earlier research, this was done by for example measuring illness acuity, length of stay or medical service use. These concepts however seem to be measuring the indirect effects of an MPU and do not directly grasp the underlying working mechanism of the MPU which is containment of disruptive behavior. Therefore, the InDiBI 3.0 instrument is a crucial addition that can be used as an outcome measurement in further research on the effectiveness of an MPU. In general, there is some resistance against instruments like the InDiBI 3.0 with only one question and a sub-question as people are wondering whether this brevity results in sufficient information. By using the comparison analysis approach of the three instruments, it turned out that this is absolutely possible. It became clear that an instrument does not have to be extensive to be valid and reliable. The findings also showed that the shortest instrument, the InDiBI 3.0, was preferred by the majority. Therefore, it is recommended to use this comparison approach in new research, as it reveals unique insights in the strengths and weaknesses of the proposed instruments. This insight is valuable for other researchers who are developing new instruments. Lastly, due to the simplicity of the InDiBI 3.0, the instrument

could also be used in other similar settings, such as psychiatric clinics and nursing homes, to evaluate the manageability of the patients and thereby the effectiveness of containment of disruptive behavior.

Afterword

Looking back on this process, I can say that I have grown a lot academically and personally. This project has been so interesting and surprised me in many ways. Firstly, I learnt that all work starts from scratch. I never really realized this, but for example the well-known instruments such as the Beck Depression Inventory or the Wechsler Adult Intelligence Scale were also developed simply through a long thought and research process; it has to start somewhere. This awareness made it possible for me to start working and developing my own instrument, together with my colleagues. My perception of research was that you need to know exactly what you want to measure and how before you can start. However, it turned out that this is not possible, as you always get confronted with obstacles along the way while you also gain new insights that lead to changes in the process. At first, I was somewhat resistant towards these constant alterations, as I am a person who prefers to plan out all steps as good as possible in advance and I was somewhat annoyed that we could not come up with a clear vision. Later on, I learnt that this was the key of research; by endless critical thinking and questioning, the research only got better and better and I am happy to have experienced this valuable process. Furthermore, I gained many new statistical skills and found out it is actually reasonably fun, when real data is used. I view the analysis part as a puzzle that has to be solved and luckily I do like puzzles. It took me many hours and I watched countless tutorial videos about analyses I had never done before, but I am proud that I learnt so much, also with the great help of my colleague. The discussion rounds with the other researchers also have been very valuable for me. They were always critical and made me push further and I learnt a lot from them. At the same time, they took my opinions and ideas seriously which gave me a lot of confidence. Conclusively, this research showed me that I am ready to continue in this academic field if I wanted to and I am very grateful that I was able to experience this feeling.

References

Bowers, L., Douzenis, A., Galeazzi, G. M., Forghieri, M., Tsopelas, C., Simpson, A., &
Allan, T. (2005). Disruptive and dangerous behavior by patients on acute psychiatric
wards in three European centres. *Social Psychiatry and Psychiatric Epidemiology*,
40(10), 822-828. doi: 10.1007/s00127-005-0967-1

Brooks, R. (2000). The reliability and validity of the Health of the Nation Outcome Scales:
validation in relation to patient derived measures. *Australian and New Zealand
Journal of Psychiatry, 34*(3), 504–511. doi:10.1046/j.1440-1614.2000.00755.x

Caarls, P. J., Van Schijndel, M. A., Van Wijngaarden, J., & Van Busschbach, J. J. (2018).
Factors influencing the admission decision for Medical Psychiatry Units: a concept
mapping approach. *Manuscript submitted for publication.*

Carmel, H., & Hunter, M. (1989). Staff Injuries From Inpatient Violence. *Psychiatric
Services*, *40*(1), 41–46. doi:10.1176/ps.40.1.41

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and
standardized assessment instruments in psychology. *Psychological Assessment, 6*(4),
284–290. doi:10.1037/100-3590.6.4.284

Clark, N., Kiyimba, F., Bowers, L., Jarrett, M., & Mcfarlane, L. (1999). 4. Absconding:
nurses views and reactions. *Journal of Psychiatric and Mental Health Nursing*, *6*(3),
219–224. doi:10.1046/j.1365-2850.1999.630219.x

Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J., & Stoddart, G. L. (2005).
Methods for the economic evaluation of health care programme. Third edition.
Oxford: Oxford University Press.

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An
Overview and Tutorial. Tutorials in *Quantitative Methods for Psychology*, 8(1), 23–
34. doi:10.20982/tqmp.08.1.p023

Hansen, M. S., Fink, P., Frydenberg , M., Oxhøj, M., Søndergaard, L., & Munk-Jørgensen, P.
(2001). Mental disorders among internal medical inpatients: Prevalence, detection, and
treatment status. *Journal of Psychosomatic Research, 50*(4), 199-204. doi:
10.1016/s0022-3999(03)00624-x

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral
sciences, Fifth edition.* Boston: Houghton Mifflin

IBM Corp. Released (2016). *IBM SPSS Statistics for Windows, Version 24.0*. Armonk, NY:
IBM Corp.

Jansen, L., van Schijndel, M., van Waarde, J., & van Busschbach, J. (2018). Health-economic outcomes in hospital patients with medical-psychiatric comorbidity: A systematic review and meta-analysis. *PLOS ONE, 13*(3), e0194029. doi:10.1371/journal.pone.0194029

Jones, T. V., Gerrity, M. S., Earp, J. (1990). Written case simulations: Do they predict physicians' behavior? *Journal of Clinical Epidemiology 43*(8), 805-815. https://doi.org/10.1016/0895-4356(90)90241-G

Kasper, J. A., Hoge, S. K., Feucht-Haviar, T., Cortina, J., & Cohen, B. (1997). Prospective study of patients' refusal of antipsychotic medication under a physician discretion review procedure. *American Journal of Psychiatry*, *154*(4), 483-489. doi:10.1176/ajp.154.4.483

Kathol, R. G. (1994). Medical psychiatry units: The wave of the future. *General Hospital Psychiatry, 16*(1), 1-3. doi:10.1016/0163-8343(94)90080-9

Kathol, R., Saravay, S. M., Lobo, A., & Ormel, J. (2006). Epidemiologic Trends and Costs of Fragmentation. *Medical Clinics of North America, 90*(4), 549–572. doi:10.1016/j.mcna.2006.04.002

Kishi, Y., & Kathol, R. G. (1999). Integrating Medical and Psychiatric Treatment in an Inpatient Medical Setting: The Type IV Program. *Psychosomatics, 40*(4), 345-355. doi: 10.1016/s0033-3182(99)71230-2

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155–163. doi:10.1016/j.jcm.2016.02.012

Leue, C., Driessen, G., Strik, J. J., Drukker, M., Stockbrügger, R. W., Kuijpers, P. M. . . . Van Os. (2010). Managing complex patients on a medical psychiatric unit: An observational study of university hospital costs associated with medical service use, length of stay, and psychiatric intervention. *Journal of Psychosomatic Research, 68*(3), 295-302. doi: 10.1016/j.jpsychores.2009.04.010

Mulder, C. L., Staring, A. B. P., Loos, J., Buwalda, V. J. A., Kuijpers, D., Sytema, S., & Wierdsma, A. I. (2004). De Health of the Nation Outcome Scales (HONOS) als instrument voor 'routine outcome assessment' [The Health of the Nation Outcome Scales (HONOS) in Dutch translation as an instrument for Routine Outcome Assessment]. *Tijdschrift voor Psychiatrie, 46*(5), 273-284.

Nijman, H. L. I., & a Campo, J. M. L. G. (2002). Situational Determinants of Inpatient Self-Harm. Suicide and Life-Threatening Behavior, *32*(2), 167–175. doi:10.1521/suli.32.2.167.24401

Rudwaleit, M., Landewe, R., van der Heijde, D., Listing, J., Brandt, J., Braun, J., … Sieper, J. (2009). The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. *Annals of the Rheumatic Diseases*, *68*(6), 770–776. doi:10.1136/ard.2009.108217

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. doi:10.1037/0033-2909.86.2.420

Silverstone, P. H. (1996). Prevalence of Psychiatric Disorders in Medical Inpatients. *The Journal of Nervous & Mental Disease, 184*(1), 43-51. doi: 10.1097/00005053-199601000-00008

Van Waarde, J. A., Richter, C., Müller, M. E. T. M., & Verwey, B. (2004). De medisch-psychiatrische unit: meerwaarde voor patiënten, artsen en ziekenhuizen. *Nederlands Tijdschrift voor Geneeskunde*, *148*(5), 209-211.

Wancata, J., Benda, N., Windhaber, J., & Nowotny, M. (2001). Does psychiatric comorbidity increase the length of stay in general hospitals? *General Hospital Psychiatry, 23*(1), 8-14. doi: 10.1016/s0163-8343(00)00110-9

Wing, J., Beevor, A., Curtis, R., Park, S., Hadden, S., & Burns, A. (1998). Health of the Nation Outcome Scales (HoNOS). Research and development. *The British Journal of Psychiatry : The Journal of Mental Science, 172*(1), 11-8. doi: 10.1192/bjp.172.1.11

Appendix A

Inpatient Disruptive Behavior Index 0.1 (InDiBI 0.1)

Werktitel "Inpatient Disruptive Behavior Index 0.1 (InDiBI 0.1)"
(Verstorend Gedrag op de Afdeling Index)

1. Agitatie / agressie* (hoNOS)
   0. Geen problemen op dit gebied
   1. Geïrriteerdheid, zoekt ruzie,
   2. Valt andere mensen lastig, bedreigt anderen, maakt kleine schade aan eigendommen, is hyperactief
   3. Fysieke agressie, dreigende houding, meer ernstige hyperactiviteit, vernieling van eigendommen
   4. Minstens één ernstige fysieke aanval op mens, vernielen van eigendommen, ernstige intimidatie

2. Zelfverwonding / suïcidaal gedrag* (hoNOS)
   0. Geen problemen op dit gebied
   1. Dreigt met verwondingen, gedachten over zelfmoord
   2. Omvat zelfverwording, bijvoorbeeld snijden
   3. Loopt ernstig lichamelijk letsel op, slaat hoofd tegen de muur
   4. Doet een zelfmoordpoging

3. Ontremming
   1. Geen problemen op dit gebied.

4. Weglopen zonder permissie
   0. Geen problemen op dit gebied
   1. Loopt af en toe weg maar komt steeds weer terug
   2. Loopt vaak weg maar komt steeds weer terug
   3. Loopt weg uit het ziekenhuis

5. Roepen, schreeuwen, geluiden maken
   0. Geen problemen op dit gebied
   1. Maakt langdurig geluiden op normaal volume
   2. Roept af en toe wat
   3. Roept vaak wat
   4. Roept de hele dag door dingen

6. Medewerking met instructies van personeel
   0. Geen problemen op dit gebied
   1. Werkt niet mee, maar gehoorzaamt uiteindelijk wel
   2. Gehoorzaamt alleen onder dwang
   3. Werkt niet mee en houdt zich niet aan de instructies

Appendix B

Inpatient Disruptive Behavior Index 0.2 (InDiBI 0.2)

Werktitel "Inpatient Disruptive Behavior Index 0.2 (InDiBI 0.2)"
(Verstorend Gedrag op de Afdeling Index)

1. Agitatie en agressie (geïnspireerd op hoNOS)
    1. Geen agitatie en agressie
    2. Geïrriteerdheid, ruzie, rusteloosheid
    3. Valt anderen lastig, bedreigt, beperkte materiele schade¸ hyperactiviteit, agitatie
    4. Fysieke agressie, dreigende houding, hyperactiviteit, vernielingen

2. Zelfverwonding / suïcidaal gedrag (geïnspireerd op hoNOS)
    1. Geen zelfverwonding en zelfmoordgedachten
    2. Dreigt met verwondingen, zelfmoordgedachten
    3. Zelfverwondingen, zelfmoordgedachten
    4. Zelfmoordpoging

3. Ontremming
    2. Geen ontremming
    3. Ongepaste opmerkingen
    4. Ongepaste opmerkingen en gedrag

4. Dwalen
    1. Dwaalt niet
    2. Dwaalt over de afdeling
    3. Loopt af en toe van de afdeling weg maar komt wel terug
    4. Loopt weg van de afdeling

5. Roepen en schreeuwen, geluiden maken
    1. Roept of schreeuwt niet
    2. Mompelt zonder aangesproken te zijn
    3. Maakt harde geluiden of roept af en toe
    4. Roept voortdurend

6. Opvolging van instructies personeel
    1. Volgt instructies op
    2. Moeizame werkrelatie, maar gehoorzaamt uiteindelijk wel
    3. Werkt niet mee en houdt zich niet aan de instructies.

Appendix C

Inpatient Disruptive Behavior Index 0.3 (InDiBI 0.3)

Inpatient Disruptive Behavior Index 0.3 (InDiBI 0.3)
(Verstorend Gedrag op de Afdeling Index)

Scoor per vraag de antwoordcategorie die het beste aansluit bij het gedrag van de patiënt.

1. Agitatie en agressie
   1. Geen agitatie en agressie
   2. Geïrriteerdheid, ruzie, rusteloosheid
   3. Valt anderen lastig, bedreigt, beperkte materiele schade¸ hyperactiviteit, agitatie
   4. Fysieke agressie, dreigende houding, hyperactiviteit, vernielingen

2. Zelfverwonding / suïcidaal gedrag
   1. Geen zelfverwonding en zelfmoordgedachten
   2. Dreigt met verwondingen, zelfmoordgedachten
   3. Zelfverwondingen, zelfmoordgedachten
   4. Zelfmoordpoging

3. Ontremming
   1. Geen ontremming
   2. Ongepaste opmerkingen
   3. Ongepaste opmerkingen en gedrag

4. Dwalen
   1. Dwaalt niet
   2. Dwaalt over de afdeling
   3. Loopt af en toe van de afdeling weg maar komt wel terug
   4. Loopt weg van de afdeling

5. Roepen en schreeuwen, geluiden maken
   1. Roept of schreeuwt niet
   2. Mompelt zonder aangesproken te zijn
   3. Maakt harde geluiden of roept af en toe
   4. Roept voortdurend

6. Opvolging van instructies personeel
   1. Volgt instructies op
   2. Moeizame werkrelatie, maar gehoorzaamt uiteindelijk wel
   3. Werkt niet mee en houdt zich niet aan de instructies

7. Onnodig om aandacht vragen
   1. Vraagt niet onnodig om aandacht
   2. Vraagt onnodig om aandacht
   3. Vraagt voortdurend onnodig om aandacht

Appendix D

Pilot study Questionnaire

Pilotstudie vragenlijst

Demografische gegevens deelnemer
Geslacht:
Leeftijd:
Functie:

Feedback
- Is de vragenlijst duidelijk?
- Zijn er categorieën die ontbreken?
- Zijn de antwoord categorieën onderscheidend?
- Andere opmerkingen?

Inpatient Disruptive Behavior Index (InDiBI)
(Verstorend Gedrag op de Afdeling Index)

Scoor per vraag de antwoordcategorie die het beste aansluit bij het gedrag van de patiënt.

1. Agitatie en agressie

   1. Geen agitatie en agressie
   2. Geïrriteerdheid, ruzie, rusteloosheid
   3. Valt anderen lastig, bedreigt, beperkte materiele schade¸ hyperactiviteit, agitatie
   4. Fysieke agressie, dreigende houding, hyperactiviteit, vernielingen

2. Zelfverwonding / suïcidaal gedrag

   1. Geen zelfverwonding en zelfmoordgedachten
   2. Dreigt met verwondingen, zelfmoordgedachten
   3. Zelfverwondingen, zelfmoordgedachten
   4. Zelfmoordpoging

3. Ontremming

   1. Geen ontremming
   2. Ongepaste opmerkingen
   3. Ongepaste opmerkingen en gedrag

4. Dwalen

   1. Dwaalt niet
   2. Dwaalt over de afdeling
   3. Loopt af en toe van de afdeling weg maar komt wel terug
   4. Loopt weg van de afdeling

5. Roepen en schreeuwen, geluiden maken

   1. Roept of schreeuwt niet
   2. Mompelt zonder aangesproken te zijn
   3. Maakt harde geluiden of roept af en toe
   4. Roept voortdurend

6. Opvolging van instructies personeel

1. Volgt instructies op
2. Moeizame werkrelatie, maar gehoorzaamt uiteindelijk wel
3. Werkt niet mee en houdt zich niet aan de instructies

7. Onnodig om aandacht vragen

1. Vraagt niet onnodig om aandacht
2. Vraagt onnodig om aandacht
3. Vraagt voortdurend onnodig om aandacht

Appendix E

Extensive results and discussion qualitative analysis pilot study InDiBI

Results

Agitation and aggression

Several respondents indicated that the difference between the answer options of this item were unclear. It was mentioned that one could be restless (*rusteloos*) without being agitated or aggressive (*geagiteerd of agressief*). To eliminate this confusion, the term restless (*rusteloos*) was left out in the improved item. This is justified as the term is unspecific and restless behavior does not always have to involve agitation or aggression, but can also be expressed by screaming, claiming behavior or wandering. Furthermore, multiple people indicated the difference in severity between threatening (*bedreigen*) and a threatening posture (*dreigende houding*) not to be logical. Someone else said agitation and aggression to be two totally different concepts with their own unique expressions. Furthermore, he did not agree that the scale was ascending as he said a conflict *(ruzie)* could be worse than hyperactivity. Also, he found bothering someone *(valt anderen lastig)* and a conflict *(ruzie)* equally bad in severity. Another doctor mentioned the difference between limited material damage (*beperkte materiële schade*) and vandalism (*vernielingen*) to be unclear, she suggested to change vandalism *(vernielingen)* into extreme vandalism (*ernstige vernielingen*). One doctor suggested to merge answer options 2 and 3 to be clearer. One doctor indicated it was unclear that hyperactivity was part of both answer option 3 and 4. She also said it might be better to split the category into two categories as there are so many options indicated in answer option 3 and 4.

The multitude of comments and the confusion about this item has led to a drastic change. This was done by simplifying the item.

Original item

1. Agitatie en agressie

    1. Geen agitatie en agressie

    2. Geïrriteerdheid, ruzie, rusteloosheid

    3. Valt anderen lastig, bedreigt, beperkte materiele schade¸ hyperactiviteit, agitatie

    4. Fysieke agressie, dreigende houding, hyperactiviteit, vernielingen

Revised item

1. Agitatie en/of agressie

    1. Geen agitatie en agressie

    2. Geïrriteerdheid en/of ruzie

    3. Valt anderen lastig, dreigende houding, beperkte materiele schade

    4. Fysieke agressie, bedreigt, hyperactiviteit, ernstige vernielingen

Self-harm and suicidal behavior

    Multiple respondents had comments about the clarity and distinctiveness of answer options 2 and 3. One doctor opted to separate self-harm (*zelfverwonding*) and suicidal behavior (*suïcidaal gedrag*), because in the current form she would not be able to score self-harm of a patient who does not have suicidal thoughts. There is no option to score aggression towards oneself. She also questioned what the difference was between answer option 2 and 3. Another nurse questioned why it stated 'threatening' (*dreigt*) in answer option 2, she suggested to change it into 'expressing' (*uiten van*). Another person misread the difference between answer option 2 and 3, as she questioned what the difference was between threatening with suicidal thoughts and suicidal thoughts. The threatening, however, is only meant to relate to the self-harm. All these different comments show that the answer options are somewhat unclear and therefore they have been adjusted. In the new version, no verb is used in answer option 2 to be more concise and minimize the chance on confusion. Answer option 3 only contains self-harm now, as this behavior can be considered more disruptive for the personnel than solely thoughts about self-harm of suicide. Furthermore, the answer option now allows it to score self-harm without suicidal thoughts having to be present.

    Other minor comments included that of a doctor who opted to change the word 'and' (*en*) into 'or' (*of*) in answer option 1, which has been adjusted. Furthermore, a nurse wanted to add the word 'visible' (*zichtbare*) in front of self-harm and she wanted to change suicidal thoughts into 'thoughts about death' (*gedachten aan de dood*). Though, these suggestions seem a bit too explicit and were therefore not granted. A more general remark was given by one doctor who indicated that this category seemed to be more related to the state of the patient than that it had something to do with disruptive behavior. This might be true in some cases, however it would still count as disruptive, as the patients would need extra attention and care of the personnel to ensure nothing disastrous happens. Hence, this category certainly fits an index to measure disruptive behavior.

Original item

2. Zelfverwonding en/of suïcidaal gedrag

    1. Geen zelfverwonding en zelfmoordgedachten

    2. Dreigt met verwondingen, zelfmoordgedachten

    3. Zelfverwondingen, zelfmoordgedachten

    4. Zelfmoordpoging

Revised item

2. Zelfverwonding en/of suïcidaal gedrag

    1. Geen (gedachten aan) zelfverwonding of zelfmoordgedachten

    2. Gedachten aan zelfverwonding en/of zelfmoordgedachten

    3. Zelfverwonding

    4. Zelfmoordpoging

Disinhibition

    One doctor indicated she was missing an answer option for inappropriate behavior only. This has been solved by making answer options 2 and 3 more distinctive. Answer option 2 only includes the inappropriate comments (ongepaste opmerkingen) now and answer option 3 only includes inappropriate behavior (ongepast gedrag) as this is more severe and disruptive . Another doctor indicated disinhibition to be more than inappropriate comments and behavior. He said inappropriate comments are often associated with sexual comments, but it is more than that. He gave some examples including eating without stopping, polydipsia or being impolite. This is certainly true, nevertheless the question does not exclude these examples. The answer options are deliberately kept broad to allow different types of disinhibited behavior to be scored. One nurse questioned whether it would be good to add something about whether the behavior would be correctable. This seems redundant .

Original item

3. Ontremming

    1. Geen ontremming

    2. Ongepaste opmerkingen

    3. Ongepaste opmerkingen en gedrag

Revised item

3. Ontremming

    1. Geen ontremming

    2. Ongepaste opmerkingen

    3. Ongepast gedrag

Wandering behavior

    Two doctors said that answer options 3 and 4 were not possible at their department, as patients have room arrest (*kamerarrest*) or because the wards are closed. Though, this will not be the case at all departments and to provide a complete category, the answer options remain part of the item. One doctor was not content with answer option 3 and suggested to change it into 'does not walk away but in need of measures to stay at the unit' (*loopt niet weg, maar heeft maatregelen nodig om op de afdeling te blijven*). He said that some patients for example have detection measures to enable this. This would also fit in to the comment another doctor made, as she said there are also patients who want to get out of bed all the time, which is not the same as wandering, but still is disruptive. Another doctor said wandering (*dwalen*) happened unconsciously and walking (*lopen*) only consciously. This additional information is not added to the question, as it does not have to make a change in being disruptive or not. The doctor also wanted to add 'and does not return' (*en komt niet terug*) to answer option 4. This has been adjusted to create a more clear and complete item.

Original item

4. Dwalen

    1. Dwaalt niet

    2. Dwaalt over de afdeling

    3. Loopt af en toe van de afdeling weg maar komt wel terug

    4. Loopt weg van de afdeling

Revised item

4. Dwalen

    1. Dwaalt niet

    2. Dwaalt over de afdeling

    3. Dwaalt en loopt weg van de afdeling maar komt wel terug

    4. Dwaalt en loopt weg van de afdeling en komt niet terug

5.

Screaming and shouting

One nurse wanted to add the word 'vocal' (*vocale*) to the sounds (*geluiden*). This seems somewhat redundant and thus is not included. All others considered the item as clear and complete. One doctor, however, questioned what was intended to be measured with this item; disinhibition or agitation or something else. This is not the case, as there are separate items to measure those dimensions. The item has been set up to measure 'noise' as a compartment of disruptive behavior.

Original item

5. Roepen en schreeuwen, geluiden maken

1. Roept of schreeuwt niet
2. Mompelt zonder aangesproken te zijn
3. Maakt harde geluiden of roept af en toe
4. Roept voortdurend

Revised item

5. Roepen en schreeuwen, geluiden maken

1. Roept en schreeuwt niet
2. Mompelt of praat zonder aangesproken te zijn
3. Maakt harde geluiden of roept af en toe
4. Roept voortdurend

Compliance with clinician instructions

Two doctors indicated that the term 'working relationship' (*werkrelatie*) was incorrect, because care givers and patients are not working colleagues. They suggested to change it into 'treatment relationship' (*behandelrelatie*), which has been adjusted as it indeed makes more sense. Another doctor mentioned the word 'difficult' (*moeizame)* to be suggestive. Another nurse wanted to add a fourth answer option 'does not cooperate consciously or unconsciously' (*werkt bewust of onbewust tegen*). Answer option 3 covers both types of this behavior, thus this item option was not added. Another nurse indicated that 'obeys' (*gehoorzaamt*) might sound denigrating and opted to change it into 'follows instructions' (*volgt instructies*). This wording has been changed.

One doctor indicated that she missed an option to score disruptive behavior which would difficult or block undergoing medical treatment, such as pulling out the infusion or other lines. She suggested this could be included in construct 6, which has been done by making the item broader. The title of the item has been changed into 'treatment cooperation' (*medewerking behandeling*) and the answer options have been adjusted to allow one to score this type of behavior as well.

Original item

6. Opvolging van instructies personeel

　　1.　Volgt instructies op

　　2.　Moeizame werkrelatie, maar gehoorzaamt uiteindelijk wel

　　3.　Werkt niet mee en houdt zich niet aan de instructies

Revised item

6. Medewerking behandeling

　　1.　Werkt mee aan behandeling en volgt instructies op

　　2.　Moeizame behandelrelatie, maar volgt instructies uiteindelijk wel op

　　3.　Werkt niet mee en houdt zich niet aan de instructies

Asking for attention unnecessarily

Almost all respondents indicated item 7 to be unclear. They said that 'asking for attention unnecessarily' (*onnodig om aandacht vragen*) was a vague and subjective term, because it is arbitrary to decide whether attention seeking is unnecessary or not. Most of the respondents suggested to use the term 'claiming behavior' (*claimend gedrag*) as they said this term is often used in hospital settings and is clearer. Therefore, this item has been completely changed.

Original item

7. Onnodig om aandacht vragen

　　1.　Vraagt niet onnodig om aandacht

　　2.　Vraagt onnodig om aandacht

　　3.　Vraagt voortdurend onnodig om aandacht

Revised item

7. Claimend gedrag

    1. Geen claimend gedrag

    2. Claimend gedrag

    3. Voortdurend claimend gedrag

General Comments

One person indicated some sentences to be very long. He suggested it might be better to use a 5-point scale for all items in ascending severity without making use of examples. This however would not suit the goals of this questionnaire to be as concise as possible. Furthermore, using a standard scale with different levels of severity would result in a loss of details, as each construct has its own specific variances in severity. Another person opted for a possibility to write down additional information about the patient's behavior.

Missing categories

Delirious behavior

Multiple respondents indicated they missed an item to score delirious behavior such as picking behavior, pulling out the infusion, catheter or other lines, or disoriented behavior. This category is indeed not covered and should be added to the index for a more complete representation of disruptive behavior.

Apathetic and catatonic patients

One doctor indicated that he missed a category to score apathetic and catatonic behavior of patients. These patients might not take much work, but they are in a bad state, which should be recorded. The doctor suggested to make a category about the interaction of the patient with the surroundings. This category is indeed not covered and should be added to the index for a more complete representation of disruptive behavior.

Random missing categories

The following categories were mentioned by different people to be missing.

- Cognition
- Intoxication
- Severity of illness
- Category about how the patient is doing

- Verbal aggression

Nevertheless, these categories do not seem to fit a particular unique type of disruptive behavior, thus will not be included. For example, a lower or higher level of cognition or being intoxicated do not always have to lead to disruptive behavior, and therefore there is no need to create an item for these elements. Severity of illness and a category about how a patient is doing also do not suit this index as there are not about behavior but purely about the physical state someone is in. Lastly, verbal aggression was mentioned to be missing, but this type of behavior can be scored in the item about agitation and aggression.

Discussion

The qualitative analysis was extremely useful in order to arrive at a desired instrument to measure the effectiveness of the MPU in terms of the burden of disruptive behavior on the personnel. The pilot study led to many new insights to improve the original version of the InDiBI. As a result, three new versions of the InDiBI were developed. The InDiBI 1.0 resembles most to the original InDiBI with a focus on severity of disruptive behavior. The InDiBI 2.0 was developed with a focus on manageability of disruptive behavior as this seemed a better indicator to measure the effectiveness of an MPU. Lastly, the InDiBI 3.0 was developed to create an even shorter instrument to measure the same construct. This version was made with the aim of delivering an concise and brief instrument in mind.

For the InDiBI 1.0, the two extra missing categories (delirious and catatonic behavior) that were mentioned in the feedback were not added in order to resemble the original version. The InDiBI 1.0 is a complete instrument, measuring the variety in disruptive behavior of the patients as viewed by the care givers.

In the process of the pilot study, new insights also led to a discussion whether the focus on severity of disruptive behavior would actually be sufficient to measure the effectiveness of an MPU. It was questioned whether this would be a good measurement instrument, as it might not be sensitive enough to detect differences among wards. A difference between the severity of disruptive behavior among a general ward and an MPU might not be detected due to the short length of stay in hospitals. Thus, with this focus, the effectiveness of an MPU in containing disruptive behavior is unlikely to be proven. Perhaps a better way to do this, is to shift the focus on how to cope with this disruptive behavior. When the care givers can cope better with the patients and when they are able to contain the disruptive behavior, more attention can be given to the somatic care, which is eventually the main goal of an MPU. Therefore, the InDiBI 2.0 was developed to score manageability of the

different types of disruptive behavior. In this version, the observers would have to score whether they could manage the disruptive behavior or not per dimension. In this version, the two missing categories 'delirious behavior' and 'catatonic behavior' were added as well.

After the development of the InDiBI 2.0, the idea to create an even shorter version of the index was brought up. This was regarded as a logical step to fulfil the aim to come up with a concise and brief instrument. Therefore, the InDiBI 3.0 consists only of two questions. The first one being 'Does the patient exhibit disruptive behavior and is this patient manageable?' and the second one being 'Which type(s) of disruptive behavior does the patient exhibit?' including 14 options to tick.

Appendix F

Inpatient Disruptive Behavior Index 1.0 (InDiBI 1.0)

Inpatient Disruptive Behavior Index 1.0 (InDiBI 1.0)

Participant nummer   _____

Instructie:

Omcirkel per gedragscategorie de ernst die het beste aansluit bij het gedrag van de patiënt.

1. Agitatie en/of agressie
    1. Geen agitatie en agressie
    2. Geïrriteerdheid en/of ruzie
    3. Valt anderen lastig, dreigende houding, beperkte materiele schade
    4. Fysieke agressie, bedreigt, hyperactiviteit, ernstige vernielingen

2. Zelfverwonding en/of suïcidaal gedrag
    1. Geen (gedachten aan) zelfverwonding of zelfmoordgedachten
    2. Gedachten aan zelfverwonding en/of zelfmoordgedachten
    3. Zelfverwonding
    4. Zelfmoordpoging

3. Ontremming
    1. Geen ontremming
    2. Ongepaste opmerkingen
    3. Ongepast gedrag

4. Dwalen
    1. Dwaalt niet
    2. Dwaalt over de afdeling
    3. Dwaalt en loopt weg van de afdeling maar komt wel terug
    4. Dwaalt en loopt weg van de afdeling en komt niet terug

5. Roepen en schreeuwen, geluiden maken
    1. Roept en schreeuwt niet
    2. Mompelt of praat zonder aangesproken te zijn
    3. Maakt harde geluiden of roept af en toe
    4. Roept voortdurend

6. Tegenwerking behandeling
    1. Werkt mee aan behandeling en volgt instructies op
    2. Moeizame behandelrelatie, maar volgt instructies uiteindelijk wel op
    3. Werkt niet mee en houdt zich niet aan de instructies

7. Claimend gedrag
    1. Geen claimend gedrag
    2. Vertoon af en toe claimend gedrag, maar is te overzien
    3. Eist alle aandacht op en vraagt steeds om hulp

Appendix G

Inpatient Disruptive Behavior Index 2.0 (InDiBI 2.0)

Inpatient Disruptive Behavior Index 2.0 (InDiBI 2.0)

Participant nummer   _____

Instructie:

Omcirkel per gedragscategorie de mate van hanteerbaarheid. Neem hierbij de meest ernstige gedragingen in gedachten die u heeft meegemaakt met de patiënt.

1. Agitatie en/of agressie
    1. Geen agitatie en agressie
    2. Agitatie en/of agressie, maar is hanteerbaar
    3. Agitatie en/of agressie en is niet hanteerbaar

2. Zelfverwonding en/of suïcidaal gedrag
    1. Geen (gedachten aan) zelfverwonding of zelfmoordgedachten
    2. Gedachten aan zelfverwonding en/of zelfmoordgedachten, maar is hanteerbaar
    3. Zelfverwonding, maar is hanteerbaar
    4. Zelfverwonding en/of zelfmoordpoging en is niet hanteerbaar

3. Ontremming (ongepaste opmerkingen en/of gedrag)
    1. Geen ontremming
    2. Ontremming, maar is hanteerbaar
    3. Ontremming en is niet hanteerbaar

4. Dwalen
    1. Dwaalt niet
    2. Dwaalt, maar is hanteerbaar
    3. Dwaalt en is niet hanteerbaar

5. Roepen en/of schreeuwen
    1. Roept en schreeuwt niet
    2. Roept en/of schreeuwt, maar is hanteerbaar
    3. Roept en/of schreeuwt en is niet hanteerbaar

6. Tegenwerking behandeling
    1. Werkt mee aan behandeling en volgt instructies op
    2. Werkt niet mee en/of houdt zich niet aan de instructies, maar is hanteerbaar
    3. Werkt niet mee en/of houdt zich niet aan de instructies en niet hanteerbaar

7. Claimend gedrag
    1. Geen claimend gedrag
    2. Claimend gedrag, maar is hanteerbaar
    3. Claimend gedrag en is niet hanteerbaar

8. Delirant gedrag
    1. Geen delirant gedrag
    2. Delirant gedrag, maar is hanteerbaar
    3. Delirant gedrag en is niet hanteerbaar

9. Catatonisch gedrag
   1. Geen catatonisch gedrag
   2. Catatonisch gedrag, maar is hanteerbaar
   3. Catatonisch gedrag en is niet hanteerbaar

Appendix H

Inpatient Disruptive Behavior Index 3 (InDiBI 3.0)

Inpatient Disruptive Behavior Index 3 (InDiBI 3.0)

Participant nummer   _____

Instructie:

Kruis aan of de patiënt verstorend gedrag vertoont en geef daarbij aan of de patiënt hanteerbaar is. Neem hierbij de meest ernstige gedragingen in gedachten die u heeft meegemaakt met de patiënt tijdens de huidige opname.

Definitie verstorend gedrag:
Gedrag van de patiënt dat u irriteert en/of vervelend vindt.

Definitie hanteerbaarheid:
Als het gedrag van de patiënt de voortgang van verpleging, diagnostiek en/of behandeling niet belemmert, is de patiënt hanteerbaar.

1. Is er sprake van verstorend gedrag door de patiënt en is de patiënt hanteerbaar?

☐   Patiënt vertoont geen verstorend gedrag en is hanteerbaar

☐   Patiënt vertoont verstorend gedrag, maar is hanteerbaar

☐   Patiënt vertoont verstorend gedrag en is niet hanteerbaar

2. Welk type verstorend gedrag vertoont de patiënt? Meerdere antwoorden mogelijk.
   ☐   Agressie
   ☐   Suïcidale uitingen
   ☐   Zelfbeschadiging
   ☐   Claimend gedrag
   ☐   Werkt niet mee met behandeling
   ☐   Wanen/achterdocht
   ☐   Hallucinaties
   ☐   Agitatie/opwinding/ plukkerig
   ☐   Ontremming of decorumverlies
   ☐   Roepen en schreeuwen
   ☐   Dwalen
   ☐   Onbeweeglijkheid
   ☐   Apathie
   ☐   Weigert te eten en/of drinken

Appendix I

Information letter for participants

Informatie voor deelname aan wetenschappelijk onderzoek

Titel onderzoek: Ontwikkeling en validatie van de Verstorend Gedrag op de Afdeling Index (*Inpatient Disruptive Behavior Index – InDiBI*): een vragenlijst voor het testen van de effectiviteit van een Medische Psychiatrische Afdeling (MPU)

Geachte heer/mevrouw,

Wij vragen u mee te doen aan een wetenschappelijk onderzoek. Meedoen is vrijwillig. Om mee te doen is wel uw schriftelijke toestemming nodig. Voordat u beslist of u wilt meedoen aan dit onderzoek, krijgt u mondelinge en schriftelijke uitleg over wat het onderzoek inhoudt. Lees deze informatie rustig door en vraag de onderzoeker uitleg als u vragen heeft.

Dit onderzoek is opgezet door onderzoekers in het Erasmus MC en wordt uitgevoerd in het Erasmus MC en Rijnstate. Voor dit onderzoek zijn 60 proefpersonen nodig.

Het doel van dit onderzoek is om de vragenlijst 'Inpatient Disruptive Behavior Index' [InDiBI, of in het Nederlands: Index voor Verstorend Gedrag bij Opgenomen Patiënten'] te valideren. Dit is de eerste vragenlijst die de hanteerbaarheid van dergelijk verstorend gedrag voor zorgprofessionals meet. Het is de bedoeling dat deze vragenlijst vanaf begin 2019 wordt gebruikt om de effectiviteit van een nieuwe Medische Psychiatrische Unit (MPU) in het Erasmus MC te meten. Onbehandelde gedragsproblemen leiden tot mindere kwaliteit van leven voor patiënten, langere opnames, meer zorgkosten en een grotere belasting voor zorgprofessionals. MPU's zijn erop gericht deze uitkomsten te verbeteren. Dit onderzoek zal er daarom aan bijdragen om de patiëntenzorg en welzijn van het personeel te verbeteren.

Uw deelname bestaat uit het beoordelen van 5 casus en een eigen patiënt door de 3 versies van de vragenlijst (InDiBI 1.0, 2 en 3) in te vullen. Daarnaast is het belangrijk dat u de formulieren 'algemene gegevens participant' en 'toestemmingsformulier' invult voor de dataverzameling- en verweking.

Bij vragen of opmerkingen kunt u contact opnemen met Eline van Oostrum (e.vanoostrum@erasmusmc.nl).

Bedankt voor uw medewerking,

E. van Oostrum, BA, student master klinische psychologie, Universiteit Leiden
Prof. Dr. J. J. van Busschbach, afdeling psychiatrie, sectiehoofd medische psychologie en psychotherapie, Erasmus MC
Drs. C. Pinto, psychiater, afdeling psychiatrie, Erasmus MC
M.A. van Schijndel MSc, psychiater, afdeling psychiatrie Rijnstate / Erasmus MC

Appendix J

Informed consent form

Toestemmingsverklaring (*informed consent*)

Titel onderzoek: Ontwikkeling en validatie van de Verstorend Gedrag op de Afdeling Index (*Inpatient Disruptive Behavior Index – InDiBI*): een vragenlijst voor het testen van de effectiviteit van een Medische Psychiatrische Afdeling (MPU)

Dit onderzoek wordt uitgevoerd door de Afdeling Psychiatrie en staat onder leiding van Prof. dr. J.J. van Busschbach, Drs. C. Pinto, psychiater en Drs. M.A. van Schijndel, psychiater

Participant nummer _____

In te vullen door de deelnemer
Ik verklaar op een voor mij duidelijke wijze te zijn ingelicht over de aard, de methode en het doel van het onderzoek. Ik weet dat de gegevens en resultaten van het onderzoek alleen anoniem en vertrouwelijk aan derden bekend gemaakt zullen worden. Mijn vragen zijn naar tevredenheid beantwoord.

Ik stem geheel vrijwillig in met deelname aan dit onderzoek. Ik behoud me daarbij het recht voor om op elk moment zonder opgaaf van redenen mijn deelname aan dit onderzoek te beëindigen.

Naam deelnemer:
_____

Datum: _____   Handtekening deelnemer:
_____

In te vullen door de uitvoerende onderzoeker
Ik heb een mondelinge en schriftelijke toelichting gegeven op het onderzoek. Ik zal resterende vragen over het onderzoek naar vermogen beantwoorden. De deelnemer zal van een eventuele voortijdige beëindiging van deelname aan dit onderzoek geen nadelige gevolgen ondervinden.

Naam onderzoeker:_____

Datum: _____   Handtekening onderzoeker: _____

Appendix K

Demographic information participants

Algemene gegevens participant

Titel onderzoek: Ontwikkeling en validatie van de Verstorend Gedrag op de Afdeling Index (*Inpatient Disruptive Behavior Index – InDiBI*) voor het testen van de effectiviteit van de medische psychiatrische afdeling

Participant nummer   _____

Geboortedatum (dd/mm/jjjj)   ____- ____- _____

Geslacht   man / vrouw / anders

Naam ziekenhuis: _____

Werkzaam op afdeling   _____

Functie   'gewoon' verpleegkundige / psychiatrisch verpleegkundige

                                                    internist / psychiater / arts-assistent

                                                    anders, nl: _____

Aantal jaren werkervaring in die functie: _____

1. Welke vragenlijst vond u het fijnste om in te vullen?

☐   InDiBI 1.0

☐   InDiBI 2.0

☐   InDiBI 3.0

Appendix L

Cases

Casus de Jonker

Meneer de Jonker is 65 jaar oud en is opgenomen in het ziekenhuis vanwege een longontsteking. Hij ligt inmiddels vier dagen op de afdeling longgeneeskunde waar hij medicatie en vocht krijgt toegediend via een infuus. Ook krijgt hij extra zuurstof via een neusbrilletje.

Vanwege een delier herkent hij niet altijd de mensen om zich heen, herinnert hij zich niet wat er is gebeurd of waarom hij in het ziekenhuis ligt. Dit maakt hem erg angstig. Hij roept voortdurend om zijn vrouw en kinderen en drukt zeker 10 keer per dag op de bel. De verpleegkundige die dan aan zijn bed komt, wordt niet zelden uitgescholden. Met veel moeite kan hij uiteindelijk wel worden gerustgesteld, maar een uur later begint de situatie opnieuw. Meneer de Jonker maakt soms seksueel getinte opmerkingen naar vrouwelijke zorgverleners.

Hij heeft ook meerdere keren zijn infuus en neusbrilletje losgetrokken en hij slaat om zich heen als deze weer vast worden gemaakt. Daarnaast probeert hij steeds uit bed te klimmen. Telkens gaat de valsensor af, en moet een verpleegkundige bij hem gaan kijken. Rustgevende medicatie is toegediend om de medische behandeling mogelijk te maken.

Casus Hassan

Mevrouw Hassan is 35 jaar oud. Zij werd opgenomen in het ziekenhuis vanwege een hypofysetumor met hoofdpijn en visusdaling. Bij dexamethason vanwege hersenoedeem krijgt zij een steroïde-geïnduceerde diabetes waarvoor frequent glucosecontrole en toediening van insuline nodig is.

Mevrouw heeft een borderline persoonlijkheidsstoornis. Ze loopt hard lachend over de afdeling, om vervolgens boos te worden en dan weer te huilen. Door haar gedrag heerst er een onrustige sfeer op afdeling. Verschillende medepatiënten noemen haar 'raar en onaangenaam'. Ze voelt zich snel bekritiseerd, flirt soms met de mannelijke medepatiënten en het personeel. Na een stevig gesprek met de seniorverpleegkundige zegt mevrouw toe haar gedrag te zullen aanpassen. Korte tijd later zegt ze tegen een aantal verpleegkundigen dat ze met hen een goede band heeft, maar dat andere verpleegkundigen (waaronder de senior) harteloos en jaloers zijn. Mevrouw vermoedt dat deze collega's haar daarom het contact met mannelijke medepatiënten en verpleegkundigen ontzeggen.

Mevrouw Hassan zegt soms dat ze dood wil, om dit bij navraag weer te ontkennen.

Casus Sardjoe

Meneer Sardjoe is 52 jaar oud en is opgenomen in het ziekenhuis na een zelfmoordpoging, waarbij hij voor de trein was gesprongen. Kort voor de poging ging zijn bedrijf failliet en vroeg zijn vrouw een echtscheiding aan. Hij zag het leven niet meer zitten. Zijn benen raakten zodanig beschadigd, dat een beenamputatie aan beide benen moest plaatsvinden.

Meneer lijdt al een jaar lang aan een zware depressie. In het ziekenhuis weigert hij medicatie, eten en drinken. Dit moet daarom worden toegediend via een infuus. Hij trekt zijn infuus er meerdere malen per dag uit, dus hij moet goed in de gaten worden gehouden. Zijn doodswens blijft actueel. Hij reageert vervolgens nauwelijks of niet op alle hulp die hij krijgt aangeboden van de zorgverleners en werkt niet mee.

Opvallend genoeg zegt hij alles na wat hem verteld wordt en maakt constant vreemde bewegingen met zijn armen. Hierdoor is het moeilijk om met de man te werken en zijn soms meerdere zorgverleners nodig om de behandeling mogelijk te maken.

Casus Steenbergen

Mevrouw Steenbergen is 24 jaar oud en is opgenomen in het ziekenhuis vanwege een urosepsis bij een verwaarloosde blaasontsteking. Zij heeft hoge koorts, een lage bloeddruk en is erg misselijk en moe. Daarnaast hebben er ook bloedingen plaatsgevonden in haar lever. Een heroïneverslaving maakt de behandeling complex.

Mevrouw heeft weinig energie en geen eetlust. Voeding en vocht moet daarom worden toegediend met een sonde en dit gaat goed. Ze slaapt slecht en heeft ook last van hallucinaties. Deze vinden vooral plaats in de nacht. Hierdoor is mevrouw erg onrustig en schreeuwt ze meerdere malen per nacht om hulp. Daarnaast heeft ze ook af en toe paniekaanvallen. Als er in deze situaties iemand haar komt helpen reageert ze vaak vijandig en boos. Het kost dan veel tijd om haar te kalmeren.

Het taalgebruik van mevrouw Steenbergen is ongepast. Ook al is zij hierop meerdere keren aangesproken, ze past haar taalgebruik niet aan. Ze vraagt steeds naar drugs of meer medicatie, omdat ze zich niet prettig voelt zonder. Ze richt zich hierbij niet alleen tot personeel, maar vraagt ook bezoekers en medepatiënten om middelen. Ze wordt boos als ze niet haar zin krijgt en vloekt dan veel.

Casus Veen

Meneer Veen is 47 jaar oud. Hij werd opgenomen vanwege een blindedarmontsteking. Na OK ligt hij nu op de verpleegafdeling.

Meneer Veen is in een slecht humeur: hij mokt en maakt cynische opmerkingen en er kan geen glimlach vanaf. Bij navraag blijkt dat hij vindt dat hij overal veel te lang op moet wachten, zoals bij het af gaan van zijn infuusalarm, wanneer hij pijn heeft en wanneer hij naar de wc moet. Bovendien smaakt het eten hem niet. Het ziekenhuis stinkt en op de afdeling is het lawaaierig. Hij werkt niet goed mee en ook de simpele dingen gaan met zuchten en steunen. Alleen met veel geduld en verdraagzaamheid lukt het om hem te verzorgen.

Meneer Veen wordt bezocht door zijn vrouw en dochter, en zij hebben zich bij het personeel en de arts herhaaldelijk beklaagd over het lange wachten, het lawaai en de stank.

Eigen patiënt

Neem de laatste patiënt die u heeft gezien in gedachten, die volgens u thuis hoorde op de Medische Psychiatrische Unit (MPU).

Eigen patiënt

Neem de laatste patiënt die u heeft gezien in gedachten, die volgens u thuis hoorde op de Medische Psychiatrische Unit (MPU).