**ONLINE SUPPLEMENTARY MATERIAL**

**Patient trust in the use of machine learning-based clinical decision support systems in psychiatric services: A randomised survey experiment**

Erik Perfalk MD1,2, Martin Bernstorff MD1,2, Andreas Aalkjær Danielsen MD PhD1,2,

Søren Dinesen Østergaard MD PhD1,2

1 Department of Affective Disorders, Aarhus University Hospital – Psychiatry, Aarhus, Denmark

2 Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

**TABLE OF CONTENT**

**PART 1: ENGLISH VERSIONS OF THE QUESTIONNAIRE AND THE ELECTRONIC INFORMATION PAMPHLETS**  
1.1. Supplementary Table 1: Baseline questionnaire

1.2. Slides from “Intervention” and “Active control”

1.2.1 Supplementary Table 2: Intervention

1.2.2 Supplementary Table 3: Active control

**PART 2: SUPPLEMENTARY TABLES AND FIGURES**

2.1 Supplementary Table 4: Clinical characteristics of the 1,098 randomized participants

2.2 PCA analysis

2.2.1 Supplementary Figure 1: Scree plot for trust items (positively worded outcome items)

2.2.2 Supplementary Table 5: PCA component loadings for trust items (positively worded outcome items)

2.2.3 Supplementary Figure 2: Scree plot for distrust items (negatively worded outcome items)

2.2.2 Supplementary Table 6: PCA component loadings for distrust items (negatively worded outcome items)

2.3 Supplementary Table 7: Single items from the post-experimental questionnaire with t-tests between groups

2.4 Supplementary Table 8: Results from intervention stratified by age, sex, diagnostic category, educational level, work status

**PART 1: ENGLISH VERSIONS OF THE QUESTIONNAIRE AND THE ELECTRONIC INFORMATION PAMPHLETS**

**1.1 Supplementary Table 1: Baseline questionnaire**

|  |  |  |
| --- | --- | --- |
|  | | |
|  | **Question/Statement** | **Options separated by commas** |
| **1** | What is your highest attained educational level? | Primary and lower secondary school, High school/Upper secondary school, Vocational school, Short-cycle higher education (less than 3 years of study),  Medium-cycle higher education (3-4 years of study),  Long-cycle higher education (more than 4 years of study), Other |
| **2** | What describes your current employment status the best? | Student (full time), Student (part time), Full-time employed, Part-time employed, Subsidized employment, Sick leave, Disability pension, Vocational rehabilitation, Unemployed, Retired, |
| **3** | How many adults are living in your household, besides yourself? | 0,1,2,3,4,5 or more |
| **4** | How many children are living in your household? | 0,1,2,3,4,5 or more, |
| **5** | Generally speaking, do you think that most people can be trusted or that you cannot be too careful in dealing with other people? | Likert scale 0-10 0: You cannot be to careful  10: Most people can be trusted |
| **6** | On a scale from 0 to 10, how much trust do you have in new technologies using computers/IT? | Likert scale 0-10 0: No trust at all  10: Full trust |
| **7** | On a scale from 0 to 10, how much trust do you have in each of the institutions listed below? |  |
| **8** | The Danish parliament | Likert scale 0-10 0: No trust at all  10: Full trust |
| **9** | The judicial system | Likert scale 0-10 0: No trust at all  10: Full trust |
| **10** | The police | Likert scale 0-10 0: No trust at all  10: Full trust |
| **11** | European Union | Likert scale 0-10 0: No trust at all  10: Full trust |
| **12** | Healthcare services | Likert scale 0-10 0: No trust at all  10: Full trust |
| **13** | Psychiatric services in Central Denmark region | Likert scale 0-10 0: No trust at all  10: Full trust |
| **14** | Would you say that you understand how decisions are made regarding your treatment in psychiatry? | Likert scale 0-10 0: No understanding at all  10: Full understanding |
| **15** | Would you say you understand what machine learning models can potentially be used for in Psychiatry? | Likert scale 0-10 0: No understanding at all  10: Full understanding |

* 1. **Slides from “Active control” and “Intervention”**

Every table cell was displayed as a single page in the electronic survey.

**1.2.1 Supplementary Table 2: Intervention**

|  |
| --- |
| **Decision processes in the Psychiatric Services with support from machine-learning models**  On the following pages, you will find a brief explanation of how machine-learning models could support decision making in the Psychiatric Services of the Central Denmark Region. It is important that you read the entire material, as it could have an impact on the study. |

|  |  |
| --- | --- |
| New technology makes it possible to use so-called machine-learning models to support decision making in the Psychiatric Services of the Central Denmark Region.  A new machine-learning model is developed by a computer being presented with material and examples that it can “learn from”.  Such material could be e.g. 10,000 electronic medical records from former patients (image 1). These records provide numerous examples of how a diagnosis is made. The computer (the machine) identifies patterns in these examples and “learns” to produce a qualified suggestion for a diagnosis based on the current patient record. This process, in which a computer learns from examples, is called machine learning (image 2).  The resulting computer programme, which can prompt a suggestion for a diagnosis based on a patient record, is called a machine-learning model (image 3). The diagnosis suggested by the machine-learning model is presented to health personnel, who can take this information into consideration. These considerations also include the staff’s own observations and results from various clinical examinations in addition to the suggestion made by the machine-learning model (images 4-5). | Et billede, der indeholder tekst, skærmbillede, design  Automatisk genereret beskrivelse |

|  |
| --- |
| A decision-making process in a doctor-patient interaction in the Psychiatric Services could play out as follows:Et billede, der indeholder computer, møbel, piano/klaver, bærbar  Automatisk genereret beskrivelse  Doctor: ”Until now, we have been treating you for an anxiety disorder. But it can be difficult to differentiate between anxiety and depression. A computer programme in our system has analysed your patient record and suggested that you might have a depression. I have thought about it, and I am inclined to agree. Therefore, I would like to ask you some questions about symptoms of depression. Would it be OK with you if we do this now?”  Patient: ”I have also considered that myself. I am ready to answer the questions.” |
| The computer programme (machine-learning model) that helped the doctor in the example above focuses on diagnostics (detection of depression). However, machine-learning models may also offer guidance in other areas of psychiatry, such as:   * Early detection of physical illness, e.g. diabetes or cardiovascular disease * Choice of optimal treatment for the individual patient (also called "personalised medicine") * Evaluation of exacerbation of depression or psychosis.   It is important to emphasise that machine-learning models can only be used for decision support when the machine-learning models have been shown to perform reliably.  It will always be the health professionals, not the machine-learning models, who will make the final decisions in psychiatry. Thus, there is a significant difference between decision support (based on the models) and the actual decision, which will always be made by health professionals.   The overall objective of using machine-learning models for decision support in psychiatry is to improve the precision and effectiveness of both the diagnosis and the treatment. |

**1.2.2 Supplementary Table 3: Active control**

|  |
| --- |
| **Decision processes in Psychiatry**  On the following pages, you will find a brief explanation of how machine-learning models could support decision making in Psychiatric Services of the Central Denmark Region. It is important that you read the entire material, as it could have an impact on the investigation. |
| In the Psychiatric Services, many decisions are made regarding the treatment of the individual patient. These decisions are based on a broad range of data (investigations, consultations, patient preferences, etc.). Therefore, these decisions are complex.  The staff’s decision-making process is often guided by information entered in the electronic patient record in connection with the patient's prior interactions with the Psychiatric Services. The illustration below shows some examples of information in the patient record.  Et billede, der indeholder tekst, skærmbillede, design  Automatisk genereret beskrivelse |
| Before a consultation with a patient, the staff prepares for the consultation based on entries in the record about the patient's prior interactions with the Psychiatric Services. A decision-making process in a doctor-patient interaction in the Psychiatric Services could play out as follows:Et billede, der indeholder computer, møbel, piano/klaver, bærbar  Automatisk genereret beskrivelse  Doctor: ”Over the past weeks, we have been treating you for an anxiety disorder. But it can be difficult to differentiate between anxiety and depression. Based on our conversations and information from your patient record, I have found that I would like to ask you some more detailed questions about symptoms of depression. Would it be OK with you if we do this now?”  Patient: ”I have also considered that myself. I am ready to answer the questions.” |
| In the above example, a joint decision is made to investigate if the patient might have a depression. Other decisions could be:   * Ordering blood samples, e.g. infection rate or blood sugar * Initiating/changing medications or psychotherapeutic treatment * Completing a course of treatment   It will vary between patients what information forms the basis of individual decisions. A common feature of all decisions is that they aim to offer the best possible care for the individual patient. |

PART 2: SUPPLEMENTARY TABLES AND FIGURES

**2.1 Supplementary Table 4. Clinical characteristics of the 1,098 randomized participants**

|  |  |  |  | **Randomization** **group** | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Overall**  n=1,098 | | **Blank Control**  n=356 | | **Active Control**  n=376 | | **Intervention**  n=366 | |
|  | **Median** | **IQR** | **Median** | **IQR** | **Median** | **IQR** | **Median** | **IQR** |
| **Age** | 34 | 25 | 33 | 25 | 35 | 25 | 34 | 25 |
| **Sex** | **n** | **%** | **n** | **%** | **n** | **%** | **n** | **%** |
| Female | 779 | 71% | 245 | 69% | 272 | 72% | 262 | 72% |
| **Most severe diagnosisa** | **n** | **%** | **n** | **%** | **n** | **%** | **n** | **%** |
| Substance use disorder (F1) | 6 | 1% | <5 | <1% | <5 | <1% | <5 | <1% |
| Psychotic disorders (F2) | 199 | 18% | 58 | 16% | 85 | 23% | 56 | 15% |
| Affective disorders (F3) | 603 | 55% | 207 | 58% | 199 | 53% | 197 | 54% |
| Neurotic disorders (F4) | 155 | 14% | 45 | 13% | 47 | 13% | 63 | 17% |
| Eating, sleeping and sexual disorders (F5) | 18 | 2% | <5 | <1% | 7 | 2% | 7 | 2% |
| Personality disorders (F6) | 41 | 4% | 14 | 4% | 16 | 4% | 11 | 3% |
| Disorders of psychological development (F8) | 9 | 1% | <5 | <1% | <5 | <1% | <5 | <1% |
| Child and adolescent disorders (F9) | 53 | 5% | 20 | 6% | 12 | 3% | 21 | 6% |
| Other | 14 | 1% | 5 | 1% | <5 | <1% | 5 | 1% |
|  | **Median** | **IQR** | **Median** | **IQR** | **Median** | **IQR** | **Median** | **IQR** |
| **Total median contacts/year** | 12 | 12 | 11 | 11 | 12 | 12 | 13 | 12 |
| **Days since first contact to Psychiatric Services** | 1,293 | 2753 | 1,179 | 2542 | 1,565 | 2902 | 1,260 | 2836 |

**a**Most severe main diagnosis during the period with available data (2011-2023). (F\*) indicates the ICD-10 chapter.

Cell counts <5 are not specified due to risk of identification of individual patients.

**2.2 Principal component analysis (PCA)**

**2.2.1 Supplementary Figure 1: Scree plot for trust items (positively worded outcome items)**

Et billede, der indeholder tekst, skærmbillede, linje/række, Kurve

Automatisk genereret beskrivelse

**2.2.2 Supplementary Table 5: PCA component loadings for trust items (positively worded outcome items)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Component loadings** | | |
| **Outcome Items (abbreviated\*)** | PC1 | PC2 | PC3 |
| Item 1: I feel safe with ML models | 0.605 | -0.362 | 0.709 |
| Item 2: Trust in ML models, | 0.604 | -0.371 | -0.705 |
| Item 7: The advantages outweigh disadvantages | 0.518 | 0.855 | -0.006 |

\*The full phrasing of the items and their scoring range are available in the methods section. PC: Principal component. ML: Machine learning.

**2.2.3 Supplementary Figure 2: Scree plot for distrust items (negatively worded outcome items)**

Et billede, der indeholder tekst, skærmbillede, linje/række, diagram

Automatisk genereret beskrivelse

**2.2.2 Supplementary Table 6: PCA component loadings for distrust items (negatively worded outcome items)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Component loadings** | | | | |
| **Outcome items (abbreviated\*)** | PC1 | PC2 | PC3 | PC4 | PC5 |
| Item 3: Malpractice due to ML | -0.452 | 0.182 | -0.268 | 0.808 | 0.195 |
| Item 4: Possibility to opt out of ML | -0.425 | -0.851 | -0.227 | -0.154 | 0.139 |
| Item 5: Dependency on ML | -0.474 | 0.131 | -0.061 | -0.106 | -0.862 |
| Item 6: Inequality in healthcare due to ML | -0.448 | 0.474 | -0.309 | -0.559 | 0.409 |
| Item 9: Errors in EHR | -0.436 | 0.011 | 0.882 | 0.002 | 0.179 |

\*The full phrasing of the items and their scoring range are available in the methods section. PC: Principal component. ML: Machine learning.EHR: Electronic health record.

**2.3 Supplementary Table 7: Single items from the post-experimental questionnaire with t-tests between groups**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Blank control vs. Active control** | | **Blank control vs. Intervention** | | **Active control vs. Intervention** | |
| **Outcome items (abbreviated\*)** | **Mean difference [CI]** | **p-value** | **Mean difference [CI]** | **p-value** | **Mean difference [CI]** | **p-value** |
| **Trust** | | | | | | |
| **I feel safe with ML models** | -0·088 [-0·517;0·342] | 0·69 | 0·503 [0·064;0·942] | 0·025 | 0·591 [0·159;1·02] | 0·0074 |
| **Trust in ML models** | 0·028 [-0·394;0·449] | 0·90 | 0·439 [0·008;0·869] | 0·046 | 0·467 [0·030;0·90] | 0·036 |
| **The advantages outweigh disadvantages of using ML models** | -0·002 [-0·412;0·408] | 0·99 | 0·466 [0·058;0·875] | 0·025 | 0·468 [0·053;0·884] | 0·027 |
| **Distrust** | | | | | | |
| **Dependency on ML models** | 0·264 [-0·212;0·740] | 0·28 | -0·114 [-0·598;0·370] | 0·64 | -0·378 [-0·869;0·103] | 0·12 |
| **Errors in EHR** | 0·014 [-0·413;0·440] | 0·95 | -0·368 [-0·809;0·073] | 0·10 | -0·382 [-0·820;0·057] | 0·088 |
| **Malpractice due to ML** | 0·054 [-0·385;0·492] | 0·81 | -0·778 [-1·22;0·333] | 0·00063 | -0·832 [-1·28;0·383] | 0·00030 |
| **Inequality in healthcare due to ML** | 0·075 [-0·396;0·546] | 0·75 | -0·359 [-0·833;0·115] | 0·14 | -0·434 [-0·902;0·034] | 0·069 |
| **Possibility to opt out ML** | -0·360 [-0·866;0·146] | 0·16 | -0·618 [-1·15;-0·086] | 0·023 | -0·258 [-0·782;0·265] | 0·33 |
| **Neutral** |  |  |  |  |  |  |
| **Explainability of a ML model** | 0·154 [-0·177;0·486] | 0·36 | 0·358 [0·017;0·698] | 0·039 | 0·203 [-0·159;0·566] | 0·27 |

\*The full phrasing of the items and their scoring range are available in the methods section. ML: Machine learning. EHR: Electronic health record. Medians (IQR) for single items are shown in table 2.

**2.4 Supplementary Table 8: Results from intervention stratified by age, sex, diagnostic category, general trust, baseline knowledge about machine learning, educational level, and current work status**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **TRUST SUM SCORE** | | | | | |
| **Stratification Variable** | **Subgroups** | **n** | **Blank control vs Active control** | **Blank control vs Intervention** | **Active control vs Intervention** |
|  |  |  | mean diff [CI] | mean diff [CI] | mean diff [CI] |
| **Sex** |  |  |  |  |  |
|  | Male | 302 | 0.951 [-1.12;3.02] | 1.05 [-1.15;3.26] | 0.101[-2.03;2.23] |
|  | Female | 690 | -0.468 [-1.76;0.827] | 1.62 [0.306;2.94] | 2.09 [0.727;3.45] |
| **Age\*** |  |  |  |  |  |
|  | <35 | 485 | -0.313[-1.89;1.26] | 1.28[-0.316;2.87] | 1.59[-0.0563; 3.24] |
|  | =>35 | 507 | 0.0195[-1.56;1.60] | 1.50[-0.130;3.14] | 1.48[-0.140;3.10] |
| **Diagnostic groups** |  |  |  |  |  |
|  | Affective/Anxiety disorders | 672 | 0.110 [-1.23;1.45] | 1.85 [0.509;3.20] | 1.74 [0.39;3.09] |
|  | Psychotic/Other | 320 | -0.497 [-2.49;1.50] | 0.364 [-1.78;2.50] | 0.860 [-1.32;3.04] |
| **General Trust\*** |  |  |  |  |  |
|  | <6 | 482 | -0.894[-2.49;0.699] | 1.39[-0.148;2.93] | 2.29[0.743;3.83] |
|  | =>6 | 510 | 0.410[-1.11;1.92] | 1.47[-0.120;3.13] | 1.06[-0.629;2.74] |
| **Baseline Knowledge of ML as CDSS\*** |  |  |  |  |  |
|  | <4 | 479 | -0.642[-2.23;0.943] | 2.11[0.518;3.69] | 2.75[1.14;4.35] |
|  | =>4 | 513 | 0.356[-1.07;1.78] | 1.19[-0.353;2.73] | 0.835[-0.712;2.38] |
| **Work status** |  |  |  |  |  |
|  | Not in work | 710 | -0.653[-1.97;0.665] | 1.07[-0.305;2.44] | 1.72[0.367;3.08] |
|  | In work | 282 | 1.29[-0.825;3.41] | 2.16[0.122;4.20] | 0.87[-1.35;3.09] |
| **Educational level** |  |  |  |  |  |
|  | Lower education | 689 | -0.225[-1.54;1.09] | 1.44[0.115;2.76] | 1.66[0.299;3.03] |
|  | Higher education | 303 | 0.172[-1.92;2.27] | 1.34[-0.883;3.56] | 1.17[-1.01;3.35] |
|  |  |  |  |  |  |
| **DISTRUST SUM SCORE** |  | | | | |
| **Stratification Variable** | **Subgroups** | **n** | **Blank control vs Active control** | **Blank control vs Intervention** | **Active control vs Intervention** |
|  |  |  | mean diff [CI] | mean diff [CI] | mean diff [CI] |
| **Sex** |  |  |  |  |  |
|  | Male | 302 | -1.31 [-4.79;2.17] | -0.27[-3.78;3.22] | 1.04[-2.65;4.73] |
|  | Female | 690 | 0.425 [-1.81;2.66] | -3.20 [-0.925;-5.48] | -3.63 [-5.89;-1.37] |
| **Age Group** |  |  |  |  |  |
|  | <35 | 485 | 0.302[-2.25;2.86] | -1.19[-3.80;1.42] | -1.49[-4.15; 1.16] |
|  | =>35 | 507 | -0.0714[-2.88;2.74] | -3.11[-5.90;-0.329] | -3.04[-5.84;-0.248] |
| **Diagnostic groups** |  |  |  |  |  |
|  | Affective/Anxiety disorders | 672 | -0.360 [-2.67;1.95] | -3.11 [-5.39;-0.823] | -2.75 [-5.07;-0.425] |
|  | Psychotic/Other | 320 | 0.861 [-2.49;4.21] | -0.232 [-3.73;3.26] | -1.09 [-4.63;2.44] |
| **General Trust** |  |  |  |  |  |
|  | <6 | 482 | -0.151[-2.88;2.57] | -1.78[-4.34;0.778] | -1.63[-4.34;1.09] |
|  | =>6 | 510 | 0.343[-2.31;3.00] | -2.75[-5.57;0.0804] | -3.09[-5.87;-0.309] |
| **Baseline Knowledge of ML as CDSS** |  |  |  |  |  |
|  | <4 | 479 | 0.146[-2.71;3.01] | -1.82[-4.53;0.89] | -1.97[-4.68;0.75] |
|  | =>4 | 513 | -0.0571[-2.55;2.44] | -3.16[-5.82;-0.491] | -3.10[-5.83;-0.363] |
| **Work status** |  |  |  |  |  |
|  | Not in work | 710 | 0.859[-1.40;3.12] | -2.13[-4.43;0.174] | -2.99[-5.24;-0.729] |
|  | In work | 282 | -1.88[-5.48;1.72] | -2.16[-5.61;1.28] | -0.283[-4.10;3.54] |
| **Educational level** |  |  |  |  |  |
|  | Lower educaion | 689 | -0.430[-2.70;1.84] | -2.12[-4.35;0.115] | -1.69[-3.97;0.596] |
|  | Higher education | 303 | 1.04[-2.44;4.53] | -2.52[-6.22;1.17] | -3.57[-7.25;0.120] |

\*Median value was used to splits into groups. CDSS: Clinical decision supoort system CI= Confidence interval. Groups for categorical variables= Diagnostic groups: Affective/anxiety disorders: ICD-10 F3-F4 chapters; Psychotic/others: ICD-10 F2, F5,F6,F8,F9, other. Work status: In work: “Employed (Full-time)”, “Employed (Part-time)”, “Student (Full-time)”, “Student (part-time)”; Not in work: “Disability Pension”, “Sick leave”, “Retired”, “Vocational rehabilitation”,”Unemployed”, “Subsidized employment”. Educational level: Higher education: “3-5 years higher education”; Lower education: “High school”, “Vocational school”, “Primary school”, “Other”.