Patient-reported effects of hospital-wide implementation of shared decision-making at a university medical centre in Germany: a pre–post trial

Fülöp Scheibler, Friedemann Geiger, Kai Wehkamp, Marion Danner, Marie Debrouwere, Constanze Stolz-Klingenberg, Anja Schuldt-Joswig, Christina Gesine Sommer, Olga Kopeleva, Claudia Bünine, Christine Wagner-Ullrich, Gerhard Koch, Marie Coors, Felix Wehking, Leonie Sundmacher, Jens Ulrich Rüffer

Abstract
Objectives To evaluate the feasibility and effectiveness of the SHARE TO CARE (S2C) programme, a complex intervention designed for hospital-wide implementation of shared decision-making (SDM).

Design Pre–post study.

Setting University Hospital Schleswig-Holstein (UKSH), Kiel Campus.

Participants Healthcare professionals as well as inpatients and outpatients from 22 departments of the Kiel Campus of UKSH.

Interventions The S2C programme is a comprehensive implementation strategy including four core modules: (1) physician training, (2) SDM support training for and support by nurses as decision coaches, (3) patient activation and (4) evidence-based patient decision aid development and integration into patient pathways. After full implementation, departments received the S2C certificate.

Main outcome measures In this paper, we report on the feasibility and effectiveness outcomes of the implementation. Feasibility was judged by the degree of implementation of the four modules of the programme. Outcome measures for effectiveness are patient-reported experience measures (PREMs). The primary outcome measure for effectiveness is the Patient Decision Making subscale of the Perceived Involvement in Care Scale (PICS<sub>DM</sub>). Pre–post comparisons were done using t-tests.

Results The implementation of the four components of the S2C programme was able to be completed in 18 of the 22 included departments within the time frame of the study. After completion of implementation, PICS<sub>DM</sub> showed a statistically significant difference (p<0.01) between the means compared with baseline. This difference corresponds to a small to medium yet clinically meaningful positive effect (Hedges’ g=0.2). Consistent with this, the secondary PREMs (Preparation for Decision Making and collaboRATE) also showed statistically significant, clinically meaningful positive effects.

Conclusions The hospital-wide implementation of SDM with the S2C-programme proved to be feasible and effective within the time frame of the programme.
HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Further studies should examine the effect of this or similar complex interventions in other healthcare settings and improve the effectiveness and resource requirements of implementation. Politically, the decisive course will now have to be set to implement the program - as originally envisaged in the Innovation Fund - across the board in German healthcare.

Introduction
Shared decision-making (SDM) is considered the apex of patient-centred, evidence-based healthcare. Much has been written about the benefits of SDM in doctor–patient interactions.1–14 Numerous intervention studies have been published reporting on the effects of different SDM interventions. Such studies have been summarised in systematic reviews at regular intervals.12 15 16 However, the implementation and scaling up of SDM remain challenging. To our knowledge, no SDM implementation has yet been demonstrated in a rigorous and quantitative evaluation to be successful and sustainable across an entire hospital or healthcare organisation.1 10 17–20

The project ‘Making SDM a Reality’ was funded from October 2017 to September 2021 (ref. 01NVF17009) by the German Innovation Fund and the Medical Faculty of Kiel University. One goal of the German Innovation Fund is to test scalability of new forms of care that have been proven effective in smaller studies. If such testing is successful, then they can be recommended for national implementation.21 22 Here, we present the feasibility and effectiveness results of the funded demonstration project.

Design and setting
The four modules of the SHARE TO CARE (S2C) programme were implemented at the entire Kiel campus of the University Hospital Schleswig-Holstein (UKSH). The Kiel campus has 25 departments, roughly 620 doctors, 1100 nurses and treats around 220 000 patients annually. This makes it 1 of the 10 largest university medical centres in Germany. The study protocol was published prospectively.1

To examine feasibility and effectiveness of the S2C programme (described in the Intervention section) we collected data from 2018 until 2021, with a baseline measurement (T₀) prior to the implementation of the S2C programme in the 20 eligible departments, and T₁ and T₂ measurements after completion of programme implementation in these departments. Departments were enrolled consecutively, with T₀ occurring between December 2018 and April 2020. Outcome measurement (for T₂) was completed in December 2021. Outcome measurement for T₁ was only reached for two departments by the end of the study (see online supplemental table 1). Data for T₁, only for neurology and neurosurgery are published elsewhere.14

Participants
We included all adult patients (age 18 and older) who in the past few weeks had an outpatient or inpatient visit to 1 of the 20 eligible departments of UKSH at the Kiel campus. Three of the 25 departments were ineligible for implementation, as all decisions with patients are made in other departments (Radiology) or were undergoing leadership changes, making participation infeasible (otolaryngology and neuropaediatrics). Two other departments were ineligible for patient-reported experience measure (PREM)-evaluation as they only treat paediatric patients (<18 years; general paediatrics and paediatric cardiosurgery). Two of the 20 eligible departments were too late for T₁ measurement due to COVID-19 delays in project schedule (orthodontics and haematology). Independent cross-sectional patient samples were included for the measurements before (T₀) and after (T₁) programme implementation. All patients who consented to research activities at their intake were included and consent was implied by the patient by returning the questionnaire.

Intervention
The multicomponent S2C programme consists of four intervention modules targeting physicians, other medical staff (nurses, medical technical assistants, physiotherapists, etc) or patients. Each module had shown effectiveness and feasibility in previously conducted randomised controlled trials.

Module 1
A minimum of 80% of all physicians of each clinical department completed a multimodal training composed of an online-training session and two consecutive individual feedback sessions based on videotaped real patient consultations. First, physicians underwent a 1-hour online training session teaching basic SDM skills, including viewing of simulated physician–patient interactions to demonstrate the dos and don’ts in SDM. Second, a real life consultation video from each physician was analysed and rated by the S2C trainer team using the MAPPING’ SDM raten manual. In small group training sessions, physicians then received individual feedback based on their own video recordings. To complete the training module, physicians recorded a second small group training and participated in a second small group training to view their, hopefully improved, performance and further develop their SDM skills.7 23 After completion physicians received their SDM certificates from the S2C programme and continuing education credits by the Physicians Chamber of Schleswig-Holstein.

Module 2
To increase patients’ participation and involvement in decision-making, every patient received information how to actively take part in their physician–patient interaction using the ‘Ask 3 Questions’ approach.13 Besides that, a comprehensive patient activation campaign across the hospital incorporated video clips (eg, on websites, bedside screens and in waiting rooms) and printed materials (flyers, roll ups, etc).

Module 3
To facilitate understanding of their health condition and available treatment options, patients received evidence-based online patient decision aids (EbPDA) in 80 different treatment decisions. To ensure content quality each EbPDA is based on a needs assessment, validated by user testing and external peer review. Clinicians at UKSH and the S2C staff collaborated to develop the list of specific decisions for which EbPDAs were developed. The evidence research team conducted a systematic review of best available evidence for all treatment opportunities available. They also
performed needs assessments with patients to align with needs and preferences of patients in the specific decision situations.

Methods of content development were based on the German Standard of Evidence-Based Patient Information and the methods of evidence generation in patient information.\textsuperscript{24–27} The process of DA development follows the International Patient Decision Aids Standards criteria.\textsuperscript{19, 20} EbPDAs should be designed to be as user-friendly as possible. Therefore, we included video sequences with UKSH physicians explaining interventions and patients sharing their experiences facing the same decision as the EbPDa users.\textsuperscript{20, 29} Patients did not report on their individual experiences with the interventions presented to avoid influence on users of the EbPDAs. An example-EbPDA can be found here: https://eh-epilepsie.share-to-care.de.

**Module 4**
All nurses were educated to help patients find relevant EbPDAs, encourage them to actively engage in decision-making, and to be aware if physicians seemed to overlook patient preferences. In addition, nurses or other healthcare professionals were trained as decision coaches to facilitate patients’ decision processes in selected clinical pathways.\textsuperscript{3, 10} Decision coaches function as emotional assistance to sensitive patients to unanswered questions and treatment preferences improving physician–patient consultation.

Training has been designed in a similar way as physicians’ face-to-face feedback sessions. However, in addition to gaining basic knowledge about SDM, decision coaches also receive detailed information about the EbPDAs of their specific department and skills to support patients’ decision-making. Nurses completed decision coach training by recording two coaching conversations with different patients at different time points and received individual feedback. Decision coaching training has an emphasis on the EbPDAs of the specific departments.

**Certification**
Once a department has met implementation and training metrics for each of the four modules, it receives a department-wide S2C certificate. Continued certification requires annual documentation by the department that the criteria has been maintained (eg, new staff have been trained).

**Data collection and outcome measures**
Detailed explanations of the study methods are published elsewhere.\textsuperscript{1} Outcome data were collected in a pre–post design via mailed patient questionnaires. Baseline measurements were conducted from July to September 2018 and immediate postintervention completion within each department (T\textsubscript{1}) from February 2020 to December 2021. Due to the COVID-19 lockdowns, implementation and evaluation were delayed for several months. This also affected the timeline and available funding for the planned evaluation. The primary intervention outcome was whether and to what degree patients were involved in clinical decisions. To cover different perspectives, we focused on two types of outcome measures, one from the patient perspective (PREM) and one from an observer perspective (Observer–Reported Outcome Measure).\textsuperscript{31} This paper focuses on the patient- perceived involvement in care (PREM). The observer-based perspective and the intermediate term outcomes (T\textsubscript{1}) are published elsewhere.\textsuperscript{32, 33} We used a validated SDM measurement instrument, the Perceived Involvement in Care Scale (PICS).\textsuperscript{34, 35} The PICS is translated and validated in Germany and consists of three subscales with 4–5 items each. The subscales are (1) Doctor Facilitation Scale (PICS\textsubscript{DFS}; five items) (2) Patient Information Scale (PICS\textsubscript{PIS}; four items) and (3) Patient Decision-making Scale (PICS\textsubscript{PDM}; five items). Each item is measured on a scale from 1 (‘do not agree at all’) to 4 (‘totally agree’). The PICS\textsubscript{PDM} subscale was prospectively defined as the primary outcome of interest.

Sampling for outcome measurement was conducted in a retrospective, consecutive index date method.\textsuperscript{1} Patients who had been discharged from the hospital before a defined date (index date) were enrolled consecutively from the index date backward until the required sample size was achieved. At T\textsubscript{1}, index dates were set for each department separately with at least 2 months gap between finalisation of interventions and start of sampling.

As the main aim of the study was to enhance patient’s perceived involvement in SDM, we focused on the (PICS\textsubscript{PDM}) subscale of the PICS.\textsuperscript{3} However, data of the other two subscales of PICS, the (PICS\textsubscript{DFS}) and the (PICS\textsubscript{PIS}), were also collected. Difference in PICS scores of 0.4 comparing T\textsubscript{0} and T\textsubscript{1} measurement (or Hedges’ g>0.5, which corresponds to a medium size effect) were used to calculate the sample size. Sample size calculation was based on expected effect size and number of returned questionnaires (attrition rates of 40% were expected).\textsuperscript{1}

Additional secondary outcomes were two validated and widely used questionnaires, the Preparation for Decision Making Scale (PrepDM: 10 items; 5-point scale; preparedness for decision-making)\textsuperscript{36} and collaboRATE (three items; 5-point scale; brief generic SDM-measure).\textsuperscript{37}

**Statistical analyses**
Data analysis was conducted by independent evaluators of the Technical University München. For descriptive purposes, data are expressed as mean with SD and/or 95% CI. Subscales were only included in the analyses if all items were filled out. Cases were excluded entirely if >95% of all items of the overall questionnaire were missing. Z-score normalisation within departments across both measurement points was used to account for heterogeneity in ‘naturally’ occurring SDM levels within each department before and after the intervention. An independent two-sided t-test was used to determine if there were significant differences between baseline and postintervention data. A score above 2.5 indicated a good level of SDM. In addition, we performed a multiple regression analysis examining the effect of age, education and gender on PICS\textsubscript{PDM}. Effect size was reported using Hedges’ g. All analyses were performed using STATA 16.1 with a p<0.05 considered statistically significant. We performed all analyses in two different data sets: data set A was the subset of all departments with at least 30 patient questionnaires returned. Data set B included the full sample of all departments as defined in the study protocol.

**Results**

**Feasibility**
As described above (see the Participants section), 20 departments were eligible for this evaluation with 18 that completed T\textsubscript{1} measurement. Eighteen of the 22 included departments have successfully implemented the full S2C programme indicating good feasibility (see online supplemental table 1 for details). Also the process evaluation gave mainly positive feedback on the different components and some suggestions for improvements.\textsuperscript{32} Fourteen departments completed the implementation and both measurements and were, therefore, suitable for the per protocol (PP) analysis.
vocational training (beyond secondary school), 34.9% had a university degree, 53.9% were retired, 63.7% were married, 83.1% had a compulsory health insurance plan and more than 95% had German as their mother tongue and had filled out the questionnaire themselves. We did not find relevant differences in these characteristics in the T₀ and T₁ subsamples (data not shown).

**Descriptive results**
Table 2 contains the analyses for PICS<sub>rom</sub>, PrepDM and collaborate-RATE for the 14 departments included in the final sample. In data set A, only those departments were included that had at least 30 questionnaires at both time points T₀ and T₁ (n=7 departaments). Data set B included all departments that had at least one questionnaire at each point in time (n=14 departments).

For PICS<sub>rom</sub> an overall difference in the mean values of 0.29 (data set A) and 0.20 (data set B) between T₀ and T₁ was shown. This difference was statistically significant and corresponded to a small to medium positive effect (Hedges' g=0.3 data set A/0.2 data set B). If, instead of the PP analysis for the primary endpoint, an intention-to-treat (ITT) analysis with including the four not fully implemented departments that also collected data at T₂, urology, cardiology, nephrology and orthopaedics/traumatology; see dashed line in online supplemental table 1), the effect remains significant but, as expected, becomes smaller (p=0.006; Hedges' g=0.13).

Consistent with this, the secondary endpoints (PrepDM and collaborate-RATE) also showed statistically significant, small but relevant positive effects (see table 2).

The responder analysis with a threshold of 2.5 on the PICS<sub>rom</sub> showed an increase in responders feeling involved in decision-making from 58.7% to 70.0% (=11.3%; data set A) and from 71.8% to 77.8% (=6.0%; data set B), respectively (see table 1).

**Subgroup analyses**
Beyond the analyses prespecified in the study protocol, explorative subgroup analyses were conducted for individual departments. However, only those who fulfilled the criteria of data set A (ie, ≥30 patients) were included in these analyses. The department with the most returned questionnaires at both time points was ophthalmology with 64 in T₀ and 80 in T₁. The departments with the fewest were neurosurgery in T₀ (n=37) and nuclear medicine in T₁ (n=37). Table 3 displays the subgroup effects for the primary endpoint (PICS<sub>rom</sub>). Similar effects were shown in the secondary endpoints (data not presented).

The PICS<sub>rom</sub> analysis of individual departments showed a significant mean increase (mean difference greater than zero, p value significant) for the departments of cardiac and vascular surgery, neurology, neurosurgery and ophthalmology. The greatest increase and the greatest effect size (Hedges' g=0.67) was shown in cardiac and vascular surgery. Of those with a significant effect size, neurology showed the smallest effect (Hedges' g=0.48, medium effect size). Effect sizes for non-significant means are not to be considered and are therefore shown in grey in table 2.

**Linear regression**
A simple linear model with the variable ‘time of measurement’ (model 1) was calculated in both data sets (see table 4). To control for sociodemographic differences (age, gender and school-leaving qualification), a linear regression of a correspondingly extended model was calculated (model 2).

The change in the coefficient for T₁ was measured in a stepwise regression with covariables (age, gender and school-leaving qualification). (model 1) was calculated in both data sets (see table 4). To control for sociodemographic differences (age, gender and school-leaving qualification), a linear regression of a correspondingly extended model was calculated (model 2).
qualification). Even after adding all the control variables, the coefficient of $T_1$ remained significant.

By adding departments with fewer than 30 returned questionnaires per measurement point (data set B), the effect (=size of $T_1$) was reduced but remained statistically significant (see table 4).

**Discussion**

In this 4-year implementation study, the complex intervention programme S2C with its four modules was fully implemented in 18 departments of UKSH, Kiel campus. This was achieved despite the fact that most departments moved to newly erected buildings during this period and that the COVID-19 pandemic repeatedly interrupted implementation efforts.

Significant and clinically meaningful effects were observed regarding patients’ perceived involvement in treatment decisions. These effects were consistent across all instruments measuring SDM with a focus on patient involvement in this project. These patient-reported results (PREMs) are also consistent with the results from the observer perspective using the MAPPIN’SDM instrument.32 As we were able to demonstrate statistically significant, positive effects on patients’ perceived involvement even after 6–18 months (ie, $T_2$) in two of the departments (neurology and neurosurgery),14 we are optimistic that these are lasting effects.

Our results suggest that SDM can be implemented in entire hospitals, across different types of departments. The observed effects are smaller than predicted in the protocol. Nevertheless, results greater than Hedges’ $g=0.2$ indicate a clinically meaningful positive effect of the intervention.

We are aware of other large-scale SDM implementation studies in Germany and other countries. The special issue of ZEFQ in 2022 summarises international developments.4 However, to our knowledge, only one of these has been evaluated in a quantitative and controlled way, comparable to this study.10 39 That study could not show an effect of the SDM intervention. The reasons for this are varied and cannot be clearly identified by directly comparing the interventions of the two studies.39 Overall, we think that for an SDM implementation to be successful, an organisation-wide cultural change is required, which is very likely to be reached only with a certain fidelity to and duration of the intervention.

**Methodological aspects**

Within the framework of this study, it was not possible to differentiate among the effects of the four individual components separately. Since all components were implemented simultaneously and interact synergistically with each other (eg, participation in the creation of the decision aids influences the clinicians’

### Table 2  SDM effect for the entire hospital (PICS, PrepDM and collaboRATE)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>PICS</th>
<th>PrepDM</th>
<th>collaboRATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data set</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n ($T_0$)</td>
<td>361</td>
<td>706</td>
<td>348</td>
</tr>
<tr>
<td>n ($T_1$)</td>
<td>440</td>
<td>581</td>
<td>414</td>
</tr>
<tr>
<td>No of departments</td>
<td>7</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>Mean ($T_0$)</td>
<td>0.16</td>
<td>0.09</td>
<td>0.12</td>
</tr>
<tr>
<td>SD ($T_0$)</td>
<td>1.02</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>Mean ($T_1$)</td>
<td>0.13</td>
<td>0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>SD ($T_1$)</td>
<td>0.96</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td>MD ($T_1$–$T_0$)</td>
<td>0.29</td>
<td>0.20</td>
<td>0.22</td>
</tr>
<tr>
<td>P value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Threshold $T_0$</td>
<td>58.7%</td>
<td>71.8%</td>
<td>n.s.</td>
</tr>
<tr>
<td>Threshold $T_1$</td>
<td>70.0%</td>
<td>77.8%</td>
<td>n.s.</td>
</tr>
<tr>
<td>Hedges’ $g$</td>
<td>0.2993</td>
<td>0.1995</td>
<td>0.2291</td>
</tr>
</tbody>
</table>

Data set A: only departments that had at least 30 valid questionnaires at both time points $T_0$ and $T_1$ were included; Data set B: all departments that had at least one valid questionnaire at each time point were included.

MD, mean difference; n.s, not specified; PICS, Perceived Involvement of Care Scale; PrepDM, preparation for decision-making.

### Table 3  SDM effect in individual departments: PICS (data set A)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>PICS</th>
<th>Cardiac and vascular surgery</th>
<th>Internal medicine</th>
<th>Neurosurgery</th>
<th>Neurology</th>
<th>Nuclear medicine</th>
<th>Ophthalmology</th>
<th>Urology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Department</td>
<td></td>
<td>General surgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n ($T_0$)</td>
<td>49</td>
<td>51</td>
<td>49</td>
<td>38</td>
<td>57</td>
<td>53</td>
<td>64</td>
<td>48</td>
</tr>
<tr>
<td>n ($T_1$)</td>
<td>66</td>
<td>62</td>
<td>51</td>
<td>74</td>
<td>70</td>
<td>37</td>
<td>80</td>
<td>70</td>
</tr>
<tr>
<td>Mean ($T_0$)</td>
<td>2.91</td>
<td>2.21</td>
<td>2.71</td>
<td>2.62</td>
<td>2.68</td>
<td>2.94</td>
<td>2.32</td>
<td>2.92</td>
</tr>
<tr>
<td>SD ($T_0$)</td>
<td>0.89</td>
<td>0.8</td>
<td>0.95</td>
<td>1</td>
<td>0.86</td>
<td>0.78</td>
<td>0.84</td>
<td>0.85</td>
</tr>
<tr>
<td>Mean ($T_1$)</td>
<td>2.8</td>
<td>2.79</td>
<td>2.91</td>
<td>3.07</td>
<td>3.05</td>
<td>2.67</td>
<td>2.75</td>
<td>2.76</td>
</tr>
<tr>
<td>SD ($T_1$)</td>
<td>0.89</td>
<td>0.91</td>
<td>0.73</td>
<td>0.81</td>
<td>0.68</td>
<td>0.95</td>
<td>0.84</td>
<td>0.91</td>
</tr>
<tr>
<td>MD ($T_1$–$T_0$)</td>
<td>0.11</td>
<td>0.58</td>
<td>0.20</td>
<td>0.45</td>
<td>0.37</td>
<td>0.27</td>
<td>0.43</td>
<td>0.16</td>
</tr>
<tr>
<td>P value</td>
<td>0.54</td>
<td>0.00</td>
<td>0.25</td>
<td>0.01</td>
<td>0.01</td>
<td>0.14</td>
<td>0.00</td>
<td>0.33</td>
</tr>
<tr>
<td>Hedges’ $g$</td>
<td>0.1148</td>
<td>0.6662</td>
<td>0.2297</td>
<td>0.5164</td>
<td>0.4758</td>
<td>0.3138</td>
<td>0.5074</td>
<td>0.1830</td>
</tr>
</tbody>
</table>

Not significant effects in grey font. Urology department=waitlist control department.

MD, mean difference; PICS, Patient Decision Making subscale of the Perceived Involvement of Care Scale; SDM, shared decision-making.
communication with the patients), it is not possible to clearly distinguish between the effects of individual SDM intervention modules. A National Institute for Health and Care Excellence (NICE) guideline on SDM yields inconsistent results regarding the desired effect size of different interventions and their combinations.40 However, there is a tendency towards intervention intensity and number of intervention components being positively correlated with effects on SDM.

Little is known about the responsiveness to change of patient-reported SDM measurement instruments.41 In our study, three PREM-instruments were used in parallel (PICS,35 PrepDM36 and collaborRATE).37 The instruments measure slightly different facets of SDM but are all suitable for a generic use across conditions and decision-making situations. The instruments provided mainly consistent results, with PrepDM and collaborRATE providing somewhat smaller effects than PICSPDM (see table 2). It should be mentioned here that the collaborRATE instrument was not used in its final German version, as this was not yet available at the time of the measurements.

The COVID-19 pandemic was likely a confounder, as it led to massive management changes in the hospital soon after the baseline survey and a significant decrease in elective procedures. While it is possible that COVID-19 biases the PREM positively (social desirability) or negatively (omission of elective interventions, increased stress of staff due to COVID measures and less time for conversations with patients), other departments that did not start SDM implementation (eg, urology) do not seem to indicate any such effect. In addition, the fact that the T2 measurements, which were largely carried out under conditions later in the pandemic, show a positive effect in two departments48 mitigates serious concerns about confounding by COVID-19.

The move of many departments to a new building during the project could have similar confounding effects. This also led to considerable additional organisational and logistical burdens. Here, we assume that it could have tended to bias the results downwards, as the intervention could have been seen as less urgent or important. Therefore, our measured effects on SDM may be an underestimate of how SDM could improve under more typical conditions.

Most studies on SDM to date have been conducted in selected patient populations who were immediately facing a preference-sensitive decision. In this study, the intervention was carried out and measured without selectively removing clinical conditions, decisions or populations. It can be assumed that a significant proportion of patients did not face a relevant decision (eg, check-up) or had an emergency admission (eg, stroke) that did not allow for SDM. In addition, many elective interventions that are typical candidates for SDM have been omitted or postponed due to the pandemic. These may also have led to an underestimate of the effect size of SDM on the PREMs we included.

**Advantages of a large-scale implementation**

It is evident that implementation of this magnitude leads to economies of scale in many ways. These include decision aid production, learning curves for trainers and processes to support implementation. For example, in the production of the 80 decision aids in this project, it was possible to improve efficiency by adjusting our processes, adding checklists and developing templates.29 While only one decision aid was completed in the first year of the project, by the end of the project we managed to complete three per month. The situation is similar for the trainings. Once an online training is in place, the individual costs for the training become lower with each additional participant. The trainers concurrently experienced a significant learning curve. At an administrative level, implementation strategies that were developed for the first departments and proved to be ineffective could be omitted in the subsequent departments. Effective implementation strategies, on the other hand, could be replicated across departments and used with only slight adjustments.

At least seven additional German hospitals will begin implementation of the S2C programme in 2023. In these hospitals, the implementation is expected to be both cheaper and faster. This is
because intervention components and implementation strategies (eg, online trainings, decision aids, patient activation materials, processes) will not need to be developed from scratch. Moreover, with each successfully implemented hospital, we hope that other hospitals will want to be seen as innovating along with their peers. Meanwhile, we also have health economic analyses that indicate that SDM can also save costs.32

Limitations
In 4 of the 20 eligible departments, the SDM interventions could not be carried out within the original project schedule. One of these, urology department did not even start with the implementation. This was due to temporary staff bottlenecks and the acute COVID-19 pandemic. However, as shown in the additional ITT analysis the overall effect stayed significant and meanwhile two of the four have also finished implementation.

This study has a pre–post design. It has neither a control group nor is it randomised. This carries the risk of several forms of bias. However, randomisation was not feasible at the level of an entire university hospital. Another possible design would have been a stepped wedge or a waiting-list control group design. The sheer size of the study, the unexpected delays in implementation schedules in the different departments and the expected heterogeneity of baseline levels of the primary outcome would also have made evaluation difficult here. According to the Medical Research Council’s scheme for evaluating complex interventions,42 43 the present project is in phase VI (long-term implementation), which is concerned with the question of whether results of interventions that proved effective in randomised trials can be replicated in non-randomised trials in the longer term or in larger scale.

Only 14 of the 20 eligible departments were included in the primary PREM analysis according to the study protocol (PP analysis). However, we additionally conducted an ITT analysis with 18 of the 20 eligible departments included, that had survey data for T1. The results stayed significant, although smaller.

Conclusion
This large-scale implementation study demonstrated the feasibility and effectiveness of implementing SDM in an entire university hospital within 4 years. Effects of individual implementation modules have been reproduced in a significantly larger way, despite a move of many departments to a new building and the COVID-19 lockdowns. Meanwhile, as a result of this project, SDM is reimbursed at the Kiel campus by all major compulsory health insurance statutory health insurance companies. In addition, the German Federal Joint Committee has issued a recommendation to make the S2C programme a national standard of care.

Author affiliations
1National Competency Center for Shared Decision Making, University Hospital Schleswig Holstein, Kiel, Germany
2Department of Paediatrics I, University Hospital Schleswig-Holstein, Kiel, Germany
3Department of Psychology, MSH Medical School Hamburg, Hamburg, Germany
4Department of Internal Medicine I, University Medical Center Schleswig-Holstein, Kiel, Germany
5Department of Health Economics, Technical University of Munich, München, Germany
6Department of Emergency Medicine, University Hospital Jena, Jena, Germany
7Center for Healthcare Organization and Implementation Research (CHOIR), Veterans Administration, Bedford, Massachusetts, USA
8Department of Population and Quantitative Health Sciences, University of Massachusetts Chan Medical School, Worcester, Massachusetts, USA
9TAKEPART Media + Science, Köln, Germany

Correction notice Since this paper was first published, an author surname has been updated to Bünzen. A second affiliation for Marie Debrouwere was also removed.

Twitter Marla Clayman @marlacleayman

Contributors FS, FG and JUR contributed to conception, design and acquisition of the study. FS, FG and JUR took responsibility for the content of the manuscript. KW, MDanner, MDebrouwere, CS-K, AS-J, CGS, OK, CW-U, GK, MC, FW and CW contributed to development and implementation of the SDM measures and design of study. FS, FG, LS, MC and CB, contributed to analysis and interpretation of data. FS, FG and JUR drafted the manuscript with all other authors having revised it critically for important intellectual content. All authors have approved the final manuscript. All authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding This work was supported by the German Innovation Fund, grant number 01NVS17009, and by the Medical Faculty of Kiel University (no grant number applicable).

Competing interests All authors received funding from the German Innovations Fund; FS, FG and JUR are co-founders of the SHARE TO CARE Patientenzentrierte Versorgung; KW received honoraria for being member of an advisory board for BMS on SDM; MC received consultancy fees from University Hospital Schleswig-Holstein LS is consultant for the Federal Ministry of Health (Germany), Federal Insurance Office and 4 statutory health insurance companies.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The Medical Ethics Committee of the Medical Faculty of Kiel University (ref. A111/18) reviewed and approved this project.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. All data relevant to the study are included in the article or uploaded as online supplemental information. Data are available at the authors on reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations,
clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Fülöp Scheibler http://orcid.org/0000-0002-4983-7253
Christina Gesine Sommer http://orcid.org/0000-0001-5446-8894
Marie Coors http://orcid.org/0000-0002-6800-3326
Felix Wehking http://orcid.org/0000-0002-7554-1851
Marla Clayman http://orcid.org/0000-0001-8491-3672

References
2 Arterburn D, Wellman R, Westbrook E, et al. Introducing decision AIDS at group health was linked to sharply lower hip and knee surgery rates and costs. Health Aff (Millwood) 2012;31:2094–104.
5 Elwyn G, Frosch DL, Kohlin S. Implementing shared decision-making: consider all the consequences. Implementation Sci 2015;11.


