

Evaluations-Framework CARE REGIO

Intention

Zweck des Evaluations-Frameworks ist es, eine einheitliche Bewertung von Systemen der digitalen Pflege zu erreichen, mit der -unter Einschränkungen- eine Bewertung verschiedener Systeme auch untereinander möglich ist. Hierzu ist es notwendig, alle möglichen Effektgrößen zu berücksichtigen sowie bei den Messungen auf quantitative direkte Messgrößen oder validierte Messinstrumente zurückzugreifen.

Untersuchungsgegenstand

Dieses Evaluationsframework ist vorgesehen für digitale Pflegetechnologien (DPT), die

1. das Pflegepersonal in der täglichen Arbeit unterstützen,
2. zur Selbstständigkeit von Pflegebedürftigen beitragen,
3. die Belastung von Angehörigen vermindern oder
4. die Ausbildung von Pflegekräften unterstützen.

Ad 1): Z.B. Entscheidungsunterstützung, Leitlinien, Verwaltungssysteme

Ad 2): Z.B. Ambulante Unterstützung im betreuten Wohnen, Robotisch unterstützte Pflege

Ad 3): Z.B. Kommunikations- und Überwachungssysteme, Früherkennungssysteme

Ad 4): Wissensbasen, eLearning-Systeme

Definitionen

Outcome:	Wirkung einer Methode
Outcome-Messung:	Quantifizierung einer Wirkung
Outcome-Werkzeug:	Werkzeug zur Outcome-Messung
Outcome-Zielgruppe:	Gruppe von Personen, auf die ein Outcome-Werkzeug angewendet wird
Outcome-Bereich	Bereich, in dem eine Wirkung auftritt (z.B. im Pflegewissen)
Wirksamkeit	Erwartete Wirkung
Effektivität	Reale Wirkung, inkl. Wirtschaftlichkeit, psychosoziale Wirkungen usw.
Kategorisierung	Vereinheitlichung der gewonnenen Messwerte zur weiteren Verarbeitung und Bildung eines Scores

Messinstrumente

In diesem Framework werden folgende Messinstrumente verwendet:

Nr	Zielbereich	Bereich	Werkzeug	Kat
1.1	Pflegerbedürftige	Physische Gesundheit	Barthel-Index	P
1.2		Psychische Gesundheit	SF12 / SF36 Short form	P
1.3a		Zufriedenheit	HCAHPS Hospital Consumer Assessment	P
1.3b			CANHELP Patient	P
1.4		Lebensqualität	WHO-QOL quality of life assessment	P
1.5		Spezifische Gesundheit	Validiertes Erhebungsinstrument z.B. Demenz	P
2.1a	Pfleger	Zufriedenheit (Beruf)	JSS Job satisfaction score	P
2.1b			CANHELP Care	P
2.2		Gesundheit	SF12 / SF36 Short form	P
2.3		Berufliche Kompetenz	CIRN Competency Inventory for Registered Nurses	P
2.4		Arbeitsbelastung	COPSOQ Copenhagen Psychosocial Questionnaire	P
3.1	Angehörige	Zufriedenheit	SWLS satisfaction of life scale	P
3.2		Psychologische Gesundheit	SF12 / SF36 Short form	P
3.3		Belastung	BSFC Burden Scale for Family Caregivers	P
4.1	Technik	Verfügbarkeit	Prozent / Jahr (inkl. Wartung)	VF
4.2		Datenqualität	Prüfung Vier-Augen-Prinzip	DQ
5.1	Organisation	Pflegequalität	PSNCQQ Satisfaction with Nursing Care Quality	P
5.2		Job-Attraktivität	Arbeitgeberattraktivität von Berthon	P
5.3		Prozesse	Performance von Prozessen	PP

Es werden zu den Zielbereichen Pflegebedürftige, Pfleger, Angehörige, Technik und Organisation jeweils Werkzeuge genannt, die verschiedenen Bereichen zugeordnet sind. Um eine Untersuchung vollständig zu gestalten, ist es sinnvoll, möglichst alle in Frage kommenden Zielbereiche (sofern sinnvoll) in möglichst vielen Bereichen zu betrachten.

Die Messwerkzeuge sind in erster Linie validierte und international etablierte Befragungen, ergänzt durch anerkannte und allgemein übliche weitere Kennzeichen.

(Spalte „Kat“ wird weiter unten erläutert.)

Messzeitpunkte

Die Messungen sollen unter normalen Umständen erst beginnen, wenn anfängliche Einschwingeffekte beendet sind. Außerdem müssen die Messungen übereinen bestimmten Mindestzeitraum erfolgen, damit periodische Fremdeffekte vermieden werden. Der Messzeitraum sollte daher frühestens 3 Monate nach Start einer Maßnahme beginnen und mindestens 6 Monate dauern.

Datenschutz

Dokumentierte Datenschutz-Zwischenfälle nach DSGVO führen zum Abbruch der Testphase. Eine erneute Evaluation sollte erst nach gründlicher Überarbeitung beginnen.

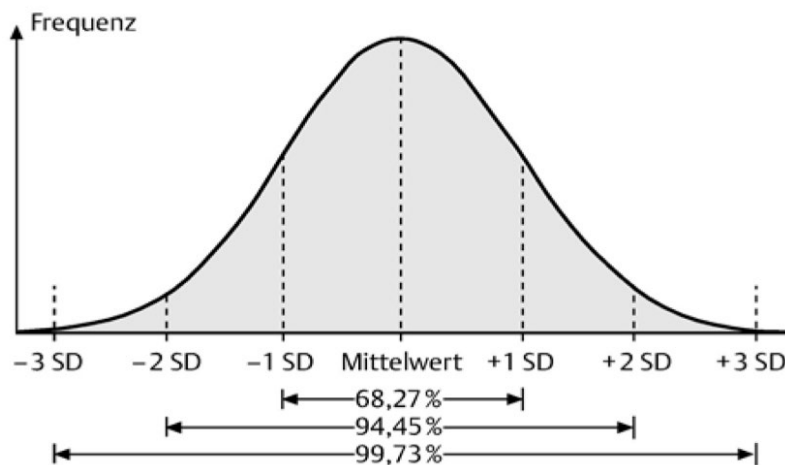
Vergleichbarkeit der Ergebnisse (Spalte „Kat“)

Hat man zwei oder mehrere neue Hilfsmittel für die Pflege zur Auswahl, dann stellt sich die Frage, welche davon die „bessere“ ist. Oder anders gesagt: Wie können wir zwei oder mehrere neue Methoden in der Pflege miteinander vergleichen?

Zunächst ist es erforderlich, einen einheitlichen Satz von Untersuchungsinstrumenten zu verwenden. Das allein ist aber nicht ausreichend, denn bei unterschiedlichen Studien ergeben sich unterschiedliche Werte, weshalb eine Standardisierung der Ergebnisse erfolgen muss. Dies geschieht in diesem Framework nach folgenden einfachen Prinzipien:

Kategorisierung nach Perzentilen bei validierten Fragebögen („Kat P“)

Das verwendete Werkzeug (in der Regel ein Fragebogen oder Score) ist validiert, d.h. es liegen Ergebnisse aus anderen Untersuchungen vor, die einer (Normal-) Verteilung folgen (siehe Abb. 1). Liegen die Ergebnisse unterhalb eines Mittelwerts, so sind die Ergebnisse schlechter als im Durchschnitt. Liegen sie darüber, sind sie besser.



Die Standardisierung erfolgt nach Vergleich der Ergebnisse mit der Verteilung des Verfahrens:

Ergebniswert x	Punkte
$x < -3 \text{ SD}$	-5
$x < -2 \text{ SD}$	-3
$x < -1 \text{ SD}$	-1
$-1 \text{ SD} < x < 1 \text{ SD}$	0
$x > 1 \text{ SD}$	1
$x > 2 \text{ SD}$	3
$x > 3 \text{ SD}$	5

Kategorisierung der technischen Verfügbarkeit („Kat VF“)

Die technische Verfügbarkeit ist ein Maß für die Zeit, in der ein technisches System dem Nutzer uneingeschränkt zur Verfügung steht.

Sie lässt sich einfach in Prozent der betriebsbereiten Zeit ausdrücken. Zum Beispiel bedeutet eine Verfügbarkeit von 99,5%, dass so ein System 99,5% der Zeit betriebsbereit ist. Andererseits ist in diesem Beispiel zu bedenken, dass eine Verfügbarkeit von 99,5%, dass dieses technische System jedes Jahr 4 Stunden und 20 Minuten nicht betriebsbereit wäre, sei es durch Ausfall oder durch Wartungsarbeiten. Die technische Verfügbarkeit sagt auch nichts darüber aus, wie oft ein System ausfällt. Im gewählten Beispiel könnte es also sein, dass das System an einem Tag 4 Stunden und 20 Minuten ausgefallen ist, genauso gut könnte es aber auch 26 mal je 10 Minuten zu unterschiedlichen Tagen ausgefallen sein. Zu den Ausfallzeiten werden auch Wartungszeiten und Ausfallzeiten einzelner, betriebswichtiger Komponenten gezählt.

Messung der Verfügbarkeit in Prozent):

$$\text{Verfügbarkeit(in Prozent)} = \left(1 - \frac{\text{Ausfallzeit}}{\text{Produktionszeit} + \text{Ausfallzeit}} \right) \cdot 100$$

Die Kategorisierung wird wie folgt vorgenommen.

Verfügbarkeitsklasse	Ausfallzeit bei 100%	Punkte
1 (<99%)	> 7:18:18 Stunden/Monat	-2
2 (99%)	7:18:18 Stunden/Monat	-1
3 (99,9%)	43:48 Minuten/Monat	0
4 (99,99%)	4:23 Minuten/Monat	1
5 (99,999%)	26,3 Sekunden/Monat oder besser	2

Kategorisierung der Datenqualität („Kat DQ“)

Gerade bei Systemen, die der medizinischen und pflegerischen Dokumentation dienen, ist nicht nur die Menge der gesammelten Daten wichtig, sondern auch deren Qualität. Beispielweise nutzt die Erhebung von 1000 Werten zu Blutdruck von Patienten wenig, wenn 10% davon falsch eingetragen wurden.

Es gibt leider nur eine valide Methode, um die Qualität von Daten zu ermitteln und das ist die stichprobenartige manuelle Kontrolle anhand der Quelldaten. Dabei sollte nach dem Vier-Augen-Prinzip vorgegangen werden, d.h. jeder Datensatz wird von zwei unabhängigen Personen überprüft. Die Stichprobengröße ist individuell festzulegen, die Ziehung der Stichprobe muss aber zufällig geschehen.

Die Kategorisierung wird wie folgt vorgenommen:

Fehlerhafte Einträge pro 1000 Datensätze	Punkte
0	2
1	1
2-5	0
6-20	-1
21-100	-2
> 100	-3

Kategorisierung der Prozessperformance („Kat PP“)

Der Einfachheit halber bezieht sich dieses Framework nur auf die zeitliche Performance von Prozessen. Werden noch andere Prozessmesswerte herangezogen, müsste das Framework entsprechend abgewandelt werden.

Hier genügt es zunächst, die zeitliche Verkürzung eines bestimmten relevanten Prozesses in Bezug auf seine Zeitdauer, ohne die neu eingesetzte Methode zu messen. Dauert die durchschnittliche pflegerische Aufnahme normalerweise 30 Minuten und kann durch Einsatz eines digitalen Assistenten diese Zeit auf 24 Minuten (durchschnittlich) verkürzt werden, so wäre ein performancegewinn von 20% zu erreichen.

Die Kategorisierung wird wie folgt vorgenommen:

Zeitliche Verbesserung	Punkte
0	0
10%	1
25%	2
> 25%	3

Ergebnisberechnung

Am Ende einer Studie wird der Mittelwert der erreichten Punkte aus den einzelnen Messwerkzeugen ermittelt und so eine Vergleichbarkeit von unterschiedlichen Studien erreicht. Zu beachten ist, dass Dauer, Stichprobengröße und Anzahl der Messwerk-

zeuge wesentlich für die Validität einer Untersuchung sind und dies beim Vergleich Berücksichtigung finden sollte.

Quellen

Krick, Tobias, Kai Huter, Dominik Domhoff, Annika Schmidt, Heinz Rothgang, und Karin Wolf-Ostermann. 2019. „Digital technology and nursing care: A scoping review on acceptance, effectiveness and efficiency studies of informal and formal care technologies“. *BMC Health Services Research* 19(1)

Swoboda W., Friedrich P., Holl F.: 15.07.2024