Report on

BCSSS-WS IV –
EVALUATION SCALES FOR DIGITAL HEALTH TECHNOLOGY ASSESSMENT

Author: Felix Tretter (BCSSS), Bertalanffy Center for the Study of Systems Science (BCSSS)
felix.tretter@bcsss.org; Mobil: 0043 660 6270 666

Thursday July 4th 2019
1040 Vienna, Paulanergasse 13 / Top 5

INTRODUCTION

The basic goal of the series of BCSSS digitization workshops is to make cyber-systemic perspectives of living systems more explicit (www.bcsss.org, 2018[i]). For this reason, health and the health care system are also important issues. The aim of the previous three workshops (WS I-III) was to explore the epistemic value of individual features of the digitization of our society, especially with regard to big data as processing technology, but also with regard to sensor and tracking technologies and analysis technologies (e.g. AI) and, in the long run, *robotics* as an epistemology technology with an increasing focus on health issues. It is therefore a question of an analytical evaluation of everything that is communicated for digitalization.

For the following, it should be noted that the more concrete but general term "Health Information and Communications Technology (Assessment)" (HICTA), which we prefer for more evident meaning, is used in the same way as the usual term "Digital Health Technology (Assessment)" (DHT or DHTA). The "feature lists" for assessment are understood as "rating scales" or "checklists" as a list of items or sub-scales or dimensions or simply questions, and the semantic background is regarded here as a "conceptual framework" and not as a "model" as is often found in the literature. The model concept is used here in a more differentiated way, whereby (theoretical) "system models" are meant to describe the structure and function of the real (empirical) system to be represented (Tretter 2005[ii]). However, even these distinctions or equations would be worth a detailed examination.

The background of the current WS IV was the final discussion at WS III, during which a 9-dimensional checklist prepared by Tretter and other physicians was discussed by HICT (DÄB, submitted). This checklist is based on a qualitative system model of health care in the sense of sustainable humane medicine (Tretter et al. 2019 [iii]). Each of these dimensions comprises approximately 5 sub-aspects or sub-scales to be tested. In this discussion, it was made clear above all that the ethical dimension must be treated separately and in detail from the legal dimension, since it is more fundamental and also affects aspects of the image of men and is not only operationally oriented towards the already very diversified legal relevance (liability law, data protection law, criminal law, duty of confidentiality, etc.). The majority of WS III also demanded that the dimension "self-reflection", i.e. the representation of the intentions of the hICT provider, should be explicitly included in the scale, since it could not be subsumed in other scales. This results in 11 dimensions with a total of estimated about 60 items.

WS IV was therefore about examining the choice of categories of the individual dimensions of the checklist more closely and comparing them with other assessment grids of the DHTA.

REPORT ON THE WORKSHOP

Workshop IV took place in her office in the NIG due to an offer of Prof. Prainsack, and was thematically and organizationally determined among the smallest circle of participants: Dr. Degelsegger-Márquez (GOEG) and Prof. Antes (Inst. f. Didaktik d. Medizin, Munich) participated, Dr. Piso (GOEG) was excused.

After Dr. Tretter's presentation of the 11-dimensional checklist mentioned above, in which he also mentioned the published system model of the health care system as a reference, and after a brief comparison with the 9-dimensional scope checklist, similarities and differences became clear: For example, the dimension "organization" in the "HTA core model", which however is depicting the complexity and dynamics of the organizational issues of HC. The "social dimension", which does not clearly distinguish between lay users and professional users of HICT, is also relatively secondary in the assessment. Ultimately, this "model" does not correspond to an explicit (e.g. social-systemic) *function model* of the health care system.

In the discussion it was pointed out that a differentiation of checklists for apps, CI, algorithms, etc. and above all between diagnostic and therapeutic tools is necessary, as it is for instance foreseen by the NICE systematics (NICE [iv]). An *ex-ante checklist* for developers of HICT would also have to be distinguished from an e*x-post checklist*. In the discussion, it was also emphasized that a differentiation of the checklists according to the intended scope of application of the HICT devices seems to make sense or vice versa: the *universal usability of the checklists* must be investigated. The wider use of existing checklists - which would also have to be partially translated - is also partly hampered by differently designed copyright regulations. An application example could be the evaluation of the electronic health card, whereby already by rough consideration the use appears very doubtful in the relationship to the organization expenditure and the negative side effects. In the absence of such German checklists, approvals have so far been made according to the criterion of “harmlessness”, which should be plausibly and comprehensibly documented. Also, these approvals by health insurance companies could aim a marketing strategy, saying “we are up to date”.

Mr. Alexander Degelsegger-Márquez - presented in his contribution different scales more exactly, like the MAST, recommendations of the EU, NICE, etc. For example, the MAST for a Multidisciplinary assessment of telemedicine distinguishes 7 scales (Kidholm et al. 2010 [v]):

1. health problem and characteristics of the application
2. safety
3. clinical effectiveness
4. patient perspectives
5. economic aspects
6. organizational aspects
7. socio-cultural, ethical and legal aspects

Criticism has been voiced (Tretter), for example, regarding the mixture of categories in so far as a distinction is made between *experimentally demonstrated efficacy* and *ecological effectiveness* as it is realized in other HTA scales (this is very important for medicine!). Also scale 7 summarizes extremely heterogeneous characteristics.

Interestingly, the “Transferability” assessment is an interesting issue of this scale: Cross-border, Scalability, Generalizability was seen as fruitful side aspects.

Ms. Barbara Prainsack proposed to set up a possible publication of the proposed checklist similar to the instrument Assert Social Impact of Security Research, which had initially been published as a

proposal on an international platform and was finally used by the EU as an assessment criterion (Dixon & Gillman 2014 [vi]). She also criticized a phenomenon (scoring) which appeared to be particularly problematic and which was particularly widespread in the USA, namely the collection of behavioral data which is actually obtained as arbitrary data in the everyday tracking of various activities and which is then suddenly attributed a predictive potential for health problems via intransparent (learning) algorithms, which is then taken up and used with interest by health insurance companies (Prainsack 2019 [vii]).

A similar problem is that data obtained through health apps are passed on to further processing instances. The processing and usage process of the data is therefore not comprehensible to the user. With the usual consent to the use of data, the further use by data analysts, which is no longer actually intended, is enabled implicitly (intransparency feature) or transparency claim.

In the further course of the event, Gerd Antes, in particular, raised a point of criticism directed at the technological faith of medicine, emphasizing that the loss of theoretical and modeling competence in medical science, which is also caused by the fixation of data, could lead to considerable misjudgments of disease mechanisms. The problem is that the formal context models, i.e. e.g. mathematical models, which are based on data analysis alone and not on subject-content theoretical components or theoretical models (e.g. variable-based behavioral equations) first have to be interpreted, which seems questionable with the excess of data and the ultimately intransparent rules of the learning algorithms (Antes 2019 [viii]). In contrast, "mechanistic models", which are based on technical expertise, are attractive in the field of health/medicine (Note Tretter: e.g. thyroid regulation models or diabetic models; see BCSSS Systems Medicine WS).

CONCLUDING REMARKS

It is necessary to have a kind of checklist on the quality of HICT available for both private and professional users in the health sector. This is entirely in the sense of consumer protection in the narrower and wider sense, especially when Microsoft in contracts of use demands agreement to exclude the possibility of legal action via consumer protection organizations? (Gigerenzer paper). The general information divide between ICT providers and users can be dangerous for HICT. In addition, there is a fundamental lack of orientation with regard to ICT processes (e.g. use of tracked data) as well as a loss of control in the case of observed dysfunctions, for example in the form of a missing "exit button" (cf. Boeing 737Max crashes). A further task is the operationalized comparison of the scales mentioned several times before in the text on the basis of a current example.

For the next workshop in fall 2019 the more detailed discussion of DHTA is planned.

SOURCES

[i] BCSSS 2019: http://www.bcsss.org/2018/big-data-epistemological-requests-of-systems-science/
[ii] Tretter, F. 2005. Systemtheorie im klinischen Kontext. Pabst, Lengerich
[iii] Tretter, F., Adam, D., Batschkus, M. 2019. . Tretter et al. Medizin in der Zange zwischen Ökonomie und Technologie. BayÄBl. 6
[iv] NICE 2018. https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/technology-appraisal-processes-guide-apr-2018.pdf
[v] Kidholm, K,, Bowes, A., Dyrehauge, S., Granstrøm-Ekeland, A., Flottorp , S.A., Jensen, L.K. Pedersen, C.D., Rasmussen, J. 2010. The MAST Manual. MAST - Model for ASsessment of Telemedicine.
[vi] Dixon, P. & Gellman, R. 'The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future' (2014) World Privacy Forum. <http://www.worldprivacyforum.org/wp-content/uploads/2014/04/WPF_Scoring_of_America_April2014_fs.pdf
[vii] Prainsack, B. (2019). Logged out: Ownership, exclusion and public value in the digital data and information commons. *Big Data & Society*: 1-15.
[viii] Antes, G. 2019: Globales Wissen, Knowledge Translation, Evidenz und Digitalisierung. https://www.cochrane.de/de/wissensakkumulation-artikel