A Consumers' Testing Approach to the Usability of Medical Technology - Insulin Pumps and CGM Systems

Tore J Larsson, Hedvig Aminoff and Mannan Mridha

School of Technology and Health
Royal Institute of Technology
100 44 Stockholm, Sweden

ABSTRACT

Five different insulin pumps and three systems for continuous glucose monitoring were subjected to usability tests at the School of Technology and Health. Each pump was trialed and rated by 30 respondents; 20 students with no experience of diabetes and 10 diabetic pump users. Each of the CGM systems was trialed and rated by 10 non-diabetic students. All participating students were enrolled in Medical Technology (Royal Institute of Technology) or Occupational Therapy (Karolinska Institute). The technical performance of pumps and CGM systems was tested independently. The respondents handled the insulin container, the software, the buttons, the screen and the manual through five scenario-based tasks. The trials and the accompanying attitude items were based on the ISO definition of usability. *Efficiency* was measured as the proportion of respondents succeeding to perform the tasks in less than 15 minutes, combined with the average time to do so. *Effectiveness* was the quotient of success frequency over average performance time. *Satisfaction w*as the average distribution on the attitude items related to software, screen, buttons and manual. All products were ranked against each other within each separate test and the rank scores accumulated. There were significant differences in the scoring of the individual insulin pumps and CGM systems.

Keywords: Usability Testing, Medical Technology, Diabetes

INTRODUCTION

The Dental and Pharmaceutical Benefits Agency, TLV, is a Swedish central government Agency whose remit is to determine if a pharmaceutical product or dental care procedure shall be subsidized by the state. This Agency has commissioned the School of Technology and Health to develop and apply a standardized evaluation model for the usability of medical technology products. Specifically, five different insulin pumps and three systems for continuous glucose monitoring available on the Swedish market were chosen for the testing.

METHOD

An evaluation model is suggested, which is simple and aimed at the ordinary consumer, defining the medical technology under scrutiny as a piece of mainstream consumer goods. The model is based on the ISO-definition of *usability*, where *efficiency, effectiveness* and *satisfaction* are measured (ISO/IEC 25062:2006; FDA 2006; NHS 2010; Jordan 1998). A number of scenario-based tasks are performed on the product by a group of respondents, who are also giving value judgements on different aspects of the product. According to this model *efficiency* was defined

https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2100-5

as the proportion of respondents who managed to finish the task with or without assistance, within 15 minutes, combined with the average time to solve the task. *Effectiveness* was defined as the quotient of success frequency over average performance time. *Satisfaction* was defined as the average distribution on the attitude items related to software, screen, buttons and manual. In addition to this, the ten queries of the System Usability Scale (Bangor et al 2008; Lewis and Sauro 2009; Sauro 2011) were put to the respondents.

Products and respondents

The five insulin pumps included in the tests were Paradigm Veo (Medtronic), Animas Vibe (Rubin Medical), Dana R (Nordic Infu Care), Accu-Chek Combo (Roche Diagnostics Scandinavia) and Omnipod (Ypsomed). The three glucose-monitoring systems were Mini Link (Medtronic), Dexcom G4 Platinum (Rubin Medical) and Freestyle Navigator II (Abbott Diabetes Care).

Each of the five pumps was tested by 20 university students (novices) and 10 diabetes patients (users). Each of the glucose monitoring systems was tested by 10 university students (novices). The students were recruited at the Royal Institute of Technology and the Karolinska Institute and were undergraduates in medical technology or occupational therapy with no prior knowledge or experience of diabetes. The diabetes patients were recruited from the large hospitals in Stockholm and Uppsala, and they were everyday users of the specific pump they were asked to test. The Omnipod pump was not available in the market; randomly selected users of other pumps tested it. The tests were conducted individually and took place in rooms at the University or the Hospital. All the pumps and CGMs were subjected to precision tests. Four different flow rates in the pumps were checked over time and the precision of the three CGM systems was compared to traditional glucose measurement.

Tasks

The respondents were given 15 minutes to study the manual and they were encouraged to set the correct time and date on the product. With the five pumps, the respondents were asked to insert the insulin capsule and prime the pump, to set a basic 24-hour program of delivery, to set a temporary 24-hour program, to set an alternative 24-hour program, and to deliver an immediate dose of insulin (bolus). With the three CGM systems the respondents were asked to set the value span for normal glucose levels, set the alarm function for highest and lowest values, and set the snooze function to repeat alarm after a certain period of time.

RESULTS

Efficiency, pumps

Efficiency was measured as frequency of success and average time to conclude each task in each pump.

Table 1: Completion rate pumps, novices (n=20)

	Task					
Completion rate	1	2	3	4	5	**average**
Accu-Chek	80	80	95	85	90	86
Animas	90	95	100	95	100	96
Dana	90	90	100	90	100	94
Medtronic	80	75	70	60	90	75
Omnipod	100	95	100	100	100	99

The Omnipod, Animas and Dana score above 90% success rate among novices; the Medtronic has the lowest success rate among novices, 75%.

Table 2: Time pumps, novices (n=20)

	Task					
Time	1	2	3	4	5	**average**
Accu-Chek	09:15	08:06	05:30	07:15	03:03	06:38
Animas	08:14	05:38	01:33	04:16	01:15	04:11
Dana	06:14	06:29	02:41	06:20	06:04	05:34
Medtronic	05:31	05:46	03:40	03:52	02:07	04:11
Omnipod	05:59	06:12	02:26	04:43	01:44	04:13

The Medtronic, Animas and Omnipod have the shortest average finishing times among novices, the Accu-Chek has the highest.

Table 3: Completion rate pumps, users (n=10)

	Task					
Completion rate	1	2	3	4	5	**average**
Accu-chek	100	90	90	80	100	92
Animas	90	100	100	80	100	94
Dana	90	80	100	89	100	92
Medtronic	100	90	80	60	90	84
Omnipod	80	100	100	100	100	96

The users scored the Medtronic pump similar to the novices; low success rate but the fastest time. The Omnipod had a high frequency of success, in spite of the fact that this was a new, unknown product to the user.

Table 4: Time pumps, users (n=10)

	Task					
Time	**1**	**2**	**3**	**4**	**5**	**average**
Accu-chek	03:32	03:42	00:50	03:31	00:20	02:23
Animas	01:40	03:02	01:37	03:52	02:01	02:26
Dana	03:37	04:14	00:41	04:57	01:10	02:56
Medtronic	01:22	02:36	00:45	02:25	00:18	01:29
Omnipod	03:07	04:56	01:56	02:56	01:02	02:47

Effectiveness, pumps

Effectiveness was recorded as completion rate/average time to finish the task. A high number represents a high degree of effectiveness.

Table 5: Effectiveness (Completion Rate Efficiency), novices (n=20)

Completion rate efficiency, novices	**1**	**2**	**3**	**4**	**5**	**average**
Accu-chek	12444	14219	24841	16901	42556	22192
Animas	15753	24284	92704	32023	115740	56101
Dana	20816	19993	53498	20451	23746	27701
Medtronic	20910	18748	27518	22337	61470	30197
Omnipod	24074	22061	59341	30482	82719	43735

Table 6: Effectiveness (Completion Rate Efficiency), users (n=10)

Completion rate efficiency, users	**1**	**2**	**3**	**4**	**5**	**average**
Accu-chek	40678	35062	155520	32817	423529	137521
Animas	77846	47446	89349	29841	71228	63142
Dana	35801	27186	209709	25848	123605	84430
Medtronic	104727	49740	153175	35711	437400	156150
Omnipod	37062	29229	74227	49203	138240	65592

Efficiency and effectiveness, CGM systems

All tasks were performed correctly, without assistance and within 10 minutes. Thus, completion rates were 100% for all three systems.

Table 7: Efficiency CGM, novices (n=10)

	Tasks		
Time	1	2	3
Dexcom	01:46	01:06	01:15
Guardian	02:58	03:46	01:55
Navigator	00:50	01:12	00:55

Performance times were shortest with the Navigator system. Completion rate efficiency is a direct mirror of performance time since completion rate was 100% for all tasks and all systems. A high number represents high effectiveness.

Table 8: Effectiveness CGM, novices (n=10)

	Tasks			
Completion rate efficiency	1	2	3	Average
Dexcom	81818	130711	115200	109243
Guardian	48458	38213	75393	54021
Navigator	173494	119834	156522	149950

Satisfaction and usability, pumps

A summed average of the responses to the six attitude items - on the pump, the program, the screen, the buttons, the manual and the insulin container – can show the proportion of respondents judging the respective aspects of the product as negative – “complicated”. Accordingly, a low number is positive.

Table 9: Satisfaction, novices (Nov) and users (Use) (n=30)

	Pump		**Software**		**Screen**		**Buttons**		**Manual**		**Container**		**∑**	**∑**
Product	No	Use	Nov	Us	Nov	Use	Nov	Use	Nov	Use	Nov	Use	Nov	Use
Accu-chek	70	30	35	10	15	30	40	10	25	22	34	30	36.5	22.0
Animas	20	10	20	10	30	0	25	0	15	30	16	10	21.0	10.0
Dana	55	20	25	30	20	20	31	11	25	22	43	10	33.2	18.8
Medtronic	45	20	11	0	11	40	11	10	17	22	0	20	15.8	18.7
Omnipod	45	20	15	0	25	20	50	20	35	60	16	22	31.0	23.7

The Animas pump had the average best scores of satisfaction as measured by novices and users combined.

The System Usability Scale (SUS, Brooke 1996) is a survey with 10 items, which represents a combined measure of users’ subjective assessment of the usability of a product or system. The result is a number between 0-100. The scale is viewed as having good validity and reliability (Bangor et al 2008). The average SUS value is 68 and this can be seen as a benchmark; a SUS value above 68 indicates better usability than the average (Sauro 2011).

Table 10: System Usability Scale, pumps (n=30)

Novices	**SUS**
Accu-chek	46
Animas	**71**
Dana	60
Medtronic	64
Omnipod	61

Users	**SUS**
Accu-chek	66
Animas	**77**
Dana	61
Medtronic	65
Omnipod	58

The highest score on usability according to SUS was recorded for the Animas pump by novices (71) and by users (77). The Accu-Chek pump attracted the lowest score from novices (46) and the Omnipod pump the lowest score from users (58).

Satisfaction and usability, CGM systems

The assessment of the CGM systems – on the five aspects of simplicity, software, screen, buttons and the manual – was averaged and ranked. Dexcom was ranked the highest, the two other systems received the same rank order.

Table 11: Satisfaction, CGM novices (n=10)

	Simplicity	Software	Screen	Buttons	Manual	Average	Rank
Dexcom	7,0	8,0	7,5	6,5	8,0	7,4	1
Guardian	5,5	5,0	4,5	7,0	3,5	5,1	2
Navigator	6,5	5,5	5,0	3,5	5,0	5,1	2

All three SGM systems received very high usability scores; a score above 80 implies extremely good usability (Sauro 2011).

Table 12: System Usability Scale, CGM systems novices (n=10)

CGM	**System Usability Scale (SUS)**	**Rank**
Dexcom	88,3	1
Guardian	80,5	3
Navigator	82,8	2

Precision, pumps

The pumps investigated were technically similar and fulfilled the specifications given in the manuals. The pumps were tested four different flow rates - 9.9, 3.3, 1.1 and 0.1 units/hour. Measurements were undertaken with the help of a high-precision scales (Sartorius BD301S) and readings were with a precision of 0.1 mg fluid. Insulin was substituted with sodium chloride (Braun 9mg/ml) and time was measured with a precision of $\pm$ 1 min.

The measurements on flow rates 9.9, 3.3 and 1.1 units/hour showed a variation of less than 5% in all of the five pumps. However, measurements of the very low flow rate 0.1 units/hour showed that three of the pumps – Animas, Dana and Omnipod – delivered in excess of the setting.

Table 13: Tested flow rates, pumps

Pump set at 9.9 units/hour	Total (mg)	Units	Measured flow units/hour	Set flow units/hour	Deviation (%)
Animas (3.12 h)	309.50	30.9	9.92	9.90	0.20
Medtronic (3.28 h)	325.20	32.5	9.91	9.90	0.15
Dana (3.12 h)	322.10	32.2	10.32	9.90	4.28
Omnipod (1.02 h)	103.60	10.3	10.16	9.90	2.59
Accu-Chek (3.15 h)	307.30	30.7	9.76	9.90	-1.46
Pump set at 3.3 units/hour					
Animas (2.02 h)	69.10	6.91	3,42	3.30	3.66
Medtronic (2.03 h)	67.20	6.72	3.31	3.30	0.31
Dana (2.03 h)	69.40	6.94	3.42	3.30	3.60
Omnipod (0.18 h)	6.20	0.62	3.44	3.30	4.38
Accu-Chek (2.02 h)	66.70	6.67	3.30	3.30	0.06
Pumps set at 1.1 units/hour					
Animas (15.70 h)	175.70	17.5	1.12	1.10	1.74
Medtronic (15.67 h)	174.40	17.4	1.11	1.10	1.18
Dana (15.67 h)	178.30	17.8	1.14	1.10	3.44
Omnipod (4.67 h)	50.60		1.08	1.10	-1.50
Accu-Chek (15.72 h)	171.90	17.1	1.09	1.10	-0.59
Pumps set at 0.1 units/hour					
Animas (29.5 h)	58.10	5.81	0.20	0.10	97.00
Medtronic (29.5 h)	30.30	3.03	0.10	0.10	3.00
Dana (29.5 h)	48.40	4.84	0.16	0.10	64.00
Omnipod (29.5 h)	37.60	3.76	0.13	0.10	27.00
Accu-Chek (29.5 h)	31.30	3.13	0.11	0.10	6.00

Precision, CGM systems

The three CGM systems, applied to a healthy member of the research team, registered measurement values over 72 hours within the target area (3.7 mmol/l to 7.8 mmol/l). The systems measured with equal precision and in agreement with glucose values taken from blood samples every 12 hours.

Estimated price pumps

The 100 novice respondents with no experience of diabetes and insulin pumps were asked to estimate the retail price of the pump they were testing. Two students responded with a very high value, the other 98 estimated the price at less than SEK 10 000. On average, the estimate is SEK 4 500:- . The Omnipod was not on the market in 2012.

Table 14: Estimated and actual retail prices insulin pumps 2012.

Pump	Estimated retail price	Actual retail price 2012
Accu-Chek	2487	30 900
Animas	8700	30 525
Dana	2783	23 000
Medtronic	2725	30 525
Omnipod	5635	-

SUMMARY OF RESULTS

The tests undertaken can be used to express four aspects of the comparative utility value of the product: *efficiency*, *effectiveness*, *satisfaction* and *usability*. Putting all the measurements together and averaging the rank orders of each of the products sums up the usability test:

Table 15: Sum of usability test, pumps (N=150)

Pumps	**Efficiency**				**Effectiveness**		**Satisfaction**		**Usability (SUS)**		**Total**
	Novice		User		Novice	User	Novice	User	Novice	User	
Accu-	1	2	3	4	1	4	1	2	1	4	**2.3**
Animas	5	4	4	3	5	1	4	5	5	5	**4.1**
Dana	2	3	3	1	2	3	2	3	2	2	**2.3**
Medtronic	5	1	1	5	3	5	5	4	4	3	**3.6**
Omnipod	3	5	5	2	4	2	3	1	3	1	**2.9**

The overall highest ranked is Animas, followed by Medtronic. The lowest ranked is Accu-Chek and Dana. The new product Omnipod is ranked in the middle.

Novices and Users differ somewhat in their judgement and performance in the tests of the pumps; Novices give Accu-Chek the lowest score, Users give the lowest score to Omnipod.

- best efficiency - Animas, Medtronic and Omnipod.
- best effectiveness - Medtronic.
- best satisfaction - Animas and Medtronic.
- best usability (SUS) - Animas.

The results are in line with the conclusions from an earlier design comparison of Animas' and Medtronics' pumps conducted at the University of Michigan (Best et al 2011).

Table 16: Sum of usability test, CGM systems, novices (N=30)

CGM	**Efficiency**		**Effectiveness**		**Satisfaction**		**Usability (SUS)**		**Total**
	Points	Rank	Points	Rank	Points	Rank	Points	Rank	Σ Rank
Dexcom	82,3	**2**	109243	**2**	1,32	**1**	88.3	**1**	**1**
Guardian	173	**3**	54021	**3**	1,17	**2**	80,5	**3**	**3**
Navigato	59	**1**	149950	**1**	0,83	**3**	82,8	**2**	**2**

The overall highest ranked is the Dexcom system, followed by Navigator and Guardian.

DISCUSSION AND CONCLUSIONS

This evaluation of insulin pumps and CGM systems shows that there are quite measurable differences in *usability*, operationalized as *efficiency*, *effectiveness* and *satisfaction*, between similar medical technology products, and that this difference can be reliably assessed with relatively uncomplicated empirical methods.

Three of the tested pumps were quite imprecise in the delivery of small quantities of insulin; this would be a problem if they were used on very small children.

The naïve estimates on pumps' values were very much lower than their actual retail prices; the underestimation must be seen in relation to respondents' experience as consumers of modern communication technology and the fact that the products tested were paid for by public funds.

A large study of attitudes among 451 young users of CGM systems identified "this was more difficult and complicated than I thought" as the strongest perceived disadvantage after 6 months of use (JDRF CGM Study Group 2008; 2009; 2010). In spite of our knowledge on the importance of *usability* and its relation to functional use and safety, problems related to the use of medical technology products remain. One reason for this is identified as the inability of manufacturers of medical technology products to understand the advantage of applying formal human factors methodology in the design process (Money et al 2011).

The problem of poorly functioning user interfaces places a heavy responsibility on the purchaser, who must choose the products best fitting the users. There are indications that decisions on purchasing often interpret usability and patient safety as technical reliability and precision (Keselman et al 2003). The design process for medical technology products needs to focus more sharply on the end user, particularly if this is the patient (Berg et al 2003; Nemeth et al 2005).

REFERENCES

Safety Management (2019)

Berg, M., Aarts, J., & van der Lei, J. (2003;42(4)). ICT in Health Care: Sociotechnical Approaches. *Methods Inf Med*, 297-301.
Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. Intl. Journal of Human–Computer Interaction, 24(6), 574-594.
Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189, 194.
FDA (1996) Do it by design: an introduction to human factors in medical devices
http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm094957.htm
http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm095061.pdf
ISO/IEC 25062:2006 Common Industry Format (CIF) for usability test reports. ISO 9241-11, ISO / IEC 9126 part 2 and 3.
Jordan, PW (1998) An Introduction to Usability. Taylor & Francis.
Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. (2008). JDRF randomized clinical trial to assess the efficacy of real-time continuous glucose monitoring in the management of type 1 diabetes: research design and methods. *Diabetes Technol Ther*, 10; 310-321.
Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. (2009). The effect of continuous glucose monitoring in well-controlled type 1 diabetes. *Diabetes Care*, 32; 1378-1383.
Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. (2010). Effectiveness of continuous glucose monitoring in a clinical care environment: evidence from the Juvenile Diabetes Research Foundation Continuous Glucose Monitoring (JDRF-CGM. *Diabetes Care* , 33, pp. 17-22.
Keselman, A., Patel, V. L., Johnson, T. R., & Zhang, J. (2003). Institutional decision-making to select patient care devices: identifying venues to promote patient safety. *Journal of biomedical informatics*, *36*(1), 31-44.
Lewis, J. R., & Sauro, J. (2009). The factor structure of the system usability scale. In Human Centered Design (pp. 94-103). Springer Berlin Heidelberg.
Money et al (2011) The role of the user within the medical device design and development process: medical device manufacturers' perspectives. *BMC Medical Informatics and Decision Making*, 11:15. http://www.biomedcentral.com/1472-6947/11/15
Nemeth, C., Nunnally, M., O'Connor, M., Klock, P., & Cook, R. (2005). Getting to the point: developing IT for the sharp end of healthcare. *J Biomed Inform , Feb;38(1)*, 18-25.
NHS (2010) Design for patient safety: user testing in the development of medical devices
http://www.nrls.npsa.nhs.uk/resources/?EntryId45=74946
Sauro, J. (2011). Measuring usability with the system usability scale (SUS). (access 2012-06-25)[2011-05-08]. Http://www.measuringusability.com/sus.php