Supplemental Information to: "Retrospective Observational Study of the Impact on Emergency Admission of Telehealth at Scale delivered in Community Care in Liverpool, UK."

Cees van Berkel[*‡],
Peter Almond[†], Carol Hughes[†],
Maurice Smith[‡], David Horsfield[‡] and Helen Duckworth[‡]

S1 Intervention length inclusion criterion

In the paper patients are included in the analysis if they were in receipt of the telehealth intervention for 9 weeks or more. The period of 9 weeks was chosen to ensure on the one hand sufficient time for the potential positive effect of the intervention to be established, while at the same time ensuring that sufficient patients would be included in the analysis for statistical confidence in the results.

Weeks>=	n	Mean	CI-Low	CI-High	p-value
1	3,889	0.07	0.04	0.1	< 0.001
2	3,819	0.07	0.04	0.1	< 0.001
4	3,740	0.07	0.04	0.1	< 0.001
6	3,669	0.07	0.04	0.1	< 0.001
9	3,562	0.08	0.05	0.11	< 0.001
12	3,429	0.08	0.05	0.11	< 0.001
16	2,644	0.08	0.05	0.11	< 0.001
20	1,581	0.05	0.01	0.09	0.02
25	842	0.05	-0.01	0.11	0.1
30	582	0.04	-0.03	0.11	0.3
40	383	0.06	-0.03	0.15	0.2
50	264	0.03	-0.09	0.15	0.6

Table S1: Treatment effect as function of number of weeks *or more* for inclusion

Weeks <	n	Mean	CI-Low	CI-High	p-value
1	94	0.01	-0.25	0.28	0.9
2	165	0	-0.19	0.19	1
4	247	-0.04	-0.2	0.12	0.6
6	316	0	-0.12	0.12	1
9	418	0	-0.1	0.09	1
12	551	0.01	-0.07	0.09	0.7
16	1,336	0.06	0.01	0.11	0.01
20	2,399	0.09	0.05	0.12	< 0.001
25	3,138	0.08	0.05	0.11	< 0.001
30	3,398	0.08	0.05	0.1	< 0.001
40	3,597	0.07	0.05	0.1	< 0.001
50	3,720	0.07	0.04	0.1	< 0.001

Table S2: Treatment effect as function of *less* than the number of weeks for inclusion

Here we provide supplemental information on the dependency of the primary outcome of that choice. Supplemental table S1 provides the number of patient/control pairs included (n), the mean *net* decrease in admissions, 95% confidence interval and p value, as function of the number-of-weeks-or-more patients had to be on telehealth to be included. In other words, the table gives the last row of table 3 in the paper for a variation of the inclusion criterion. The 6th row (9 weeks or more) reproduces the numbers in the paper. As can be seen for weeks<9, n is slightly larger and the mean effect slightly less, because of the inclusions of early dropouts.

The mean *net* decrease in admissions in the first row is 0.07, but this does not mean that it is possible to say with confidence that patients who were on telehealth for just a week enjoyed a positive benefit. Table S2 provides complementary information to this. It gives results as function of the number-of-weeks-*or-less* that the patients had to be on service to be included.

[*]Philips Research UK. Corresponding author, current address: Cees.van.berkel@hotmail.co.uk
[†]Liverpool Community Health (LCH)
[‡]Liverpool Clinical Commissioning Group (CCG)

van Berkel C, *et al. BMJ Open* 2019; 9:e028981. doi: 10.1136/bmjopen-2019-028981

Tables S1 and S2 also suggest, but do not prove, that patients who were on service for very long, experience a slightly smaller benefit. Remembering that the decision to say on telehealth for longer than the recommended 3 months is made in consultation with patient, there may be complex reasons for this effect, if real.

S2 Effect of the evaluation window

From month	to month	n	Mean	CI-Low	CI-High	p-value
0	12	3,562	0.08	0.05	0.11	< 0.001
3	15	3,102	0.09	0.06	0.12	< 0.001
6	18	2,768	0.07	0.04	0.11	< 0.001
12	24	2,118	0.06	0.02	0.1	0.002
18	30	1,618	0.02	-0.02	0.07	0.3
24	36	1,122	0.05	0	0.11	0.05
30	42	529	0.04	-0.03	0.12	0.3
36	48	70	0.11	-0.14	0.36	0.4

Table S3: Treatment effect as function of time window after start

The paper provides a before/after comparison of closely matched intervention/control pairs. The 'before' time window in the paper is 12 months prior to the start of service, while the 'after' window is 12 months from the start of telehealth. The width of this window was fixed by the data, but it is possible to shift the 'after' window in time in order to gain some insight into the retention of the effect of telehealth. Table S3 provides the number of patient/control pairs included (n), the mean *net* decrease in admissions, 95% confidence interval and p value, as function of the 'after' time window. As can be seen the positive effect is retained for some considerable time.

S3 Propensity Score

Description	mean	r	or	p
Average number of vital signs (Weight, BP, Heart rate, Glucose) submitted per day (continuous number 0-5)	1.85	0.036	1.046	0.03
Daily rate at which a patient complied with telehealth tasks (continuous number, 0 is never, 1 is daily)	0.41	0.044	0.96	0.56
Patient reported a decrease in GP or hospital visits (Yes/No)	0.49	-0.003	0.993	0.82
Patient reported more control, confidence or ability to cope (Yes/No)	0.83	-0.018	1.03	0.48
Patient reported lifestyle (diet, exercise) improvement (Yes/No)	0.43	0.005	1.014	0.66
Patient reported sharing of telehealth results with others (Yes/No)	0.71	-0.015	0.966	0.32
Patient reported improved health or better health management (Yes/No)	0.7	0.014	1.01	0.78
Patient recommended telehealth to friends or family (Yes/No)	0.48	-0.013	0.97	0.34
Patient reported Willingness to use Motiva in the future (Yes/No)	0.73	-0.03	0.951	0.17

Table S4: Engagement covariates, mean=population mean; r= univariate spearman correlation; or= univariate odds ratio, p= significance level

The paper uses a propensity score that combines the risk for emergency admissions and predictive telehealth activity indicators to predict an outcome of positive benefit of telehealth.

Engagement data was available for 2,557 patients who also satisfy the three inclusion criteria. No engagement data was available for controls and matching was not done on engagement parameters. The engagement data were made up of 9 co-variates for a predictive model, 2 were measures of telehealth compliance and 7 were self-reported responses to a general engagement questionnaire.

The TRIPOD checklist[1] is a useful tool for development and reporting of prospective scoring. Although the score wat not part of the primary outcome of the paper, the checklist has been completed and uploaded additional supplemental information to demonstrate compliance with the key items of the checklist and the TRIPOD statement itself[2].

Table S4 provides a summary of the engagement questionnaire. Correlation, odds ratios and significance levels with respect to the primary outcome are also shown in the table. It can be seen that these are weak. A general linear model is built to predict the primary outcome from these co-variates. An 80/20 split is made in the data for training and validation and repeated 5 times to cross validate predictions for all 2,557 data points.

Because the model is relatively weak, it is combined 50/50 with the calculated risk at start of service. That is an equal weight combination of the regression score of the telehealth activity indicators and Emergency Admissions risk calculated using the Welsh Model [12] parameters, the latter with regression coefficients that were optimized for the Liverpool population.

To evaluate this model, the Lorentz curve and Gini coefficient are used. These are well known measures of inequality[3]. They are parallel measures to ROC curve and AUC, but applicable to a larger class of predictive

[1] https://www.tripod-statement.org/

[2] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med. 2015;162:W1-W73. doi: 10.7326/M14-0698

[3] Cowell, F. A. (1995) Measuring Inequality Harvester Wheatshef: Prentice Hall.

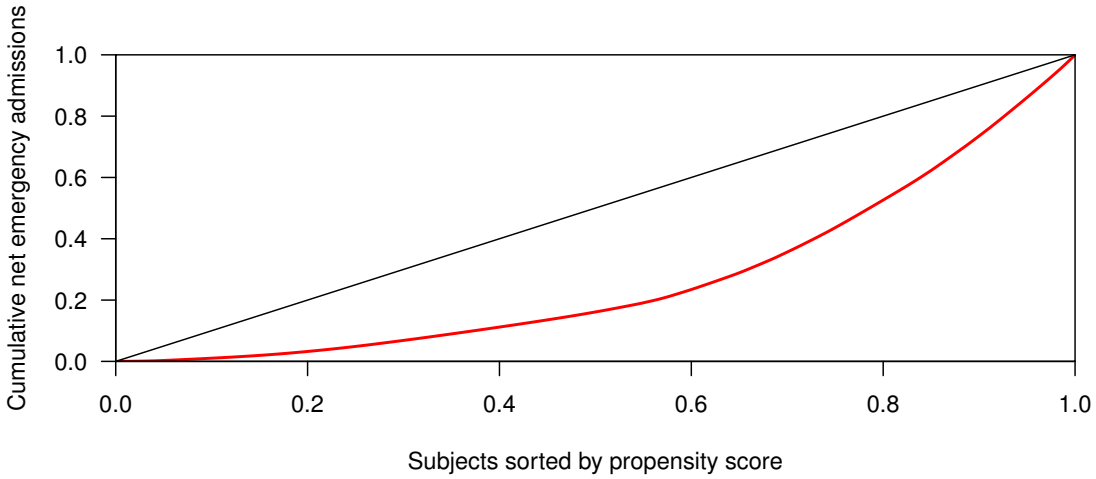

Figure S1: Lorentz curve for the propensity score

models. As such they are gaining use in data science[4]. Intuitively this makes sense because good predictive models show high 'inequality' of the actual observed outcome with respect to a predicted outcome. The Gini coefficient is a number between 0 and 1 and a high value indicates a good predictive model. For binary classifiers the AUC (A) and Gini coefficient (G) are related by

$$G = 2A - 1$$

For the model here, figure S1 gives the Lorentz curve for the propensity model. Similar to the ROC the diagonal line represents a poor, random choice model. An ideal model is one in which the Lorenz curve tucks perfectly into the bottom right corner. The Gini coefficient is the fraction that the area between the diagonal line and the Lorentz curve represents of the total area under the diagonal line. The Gini coefficient in this case is 0.46, indicating a moderately good model.

Using this model, cut-off points can be chosen to see if a patient group can be identified that benefits more from the intervenion than the cohort as a whole. Table S5 illustrates the primary outcome of the paper if the intervention group is limited to a subset patients with high prospective score. As can be seen, patients with a high prospective score benefit, on average, more than the the cohort as a whole. Note that the bottom row of this table is slightly different from the bottom row of table 3 in the paper because the result here could only be presented for patients with engagement data.

Subset	n	Mean	CI-Low	CI-High	p-value
25%	639	0.16	0.07	0.25	6e-04
50%	1,278	0.15	0.09	0.2	< 0.001
75%	1,917	0.1	0.06	0.14	< 0.001
100%	2,557	0.09	0.05	0.12	< 0.001

Table S5: Treatment effect as function top slice of patients with high prospective score

Referring to the second row in table S5 statistically significant decrease exists for the patients with an above median propensity score of 0.15, 95% confidence interval 0.09 to 0.2, corresponding to a percentage decrease in admissions of 25.9%.

[4]For instance: https://www.kaggle.com/c/porto-seguro-safe-driver-prediction