Developing a Method to Derive Indicative Health Literacy from Routine Socio-Demographic Data

Karin R Laursen[1,2*]**,**
Paul T Seed[3]**,**
Joanne Protheroe[4]**,**
Michael S Wolf[5]**,**
Gillian P Rowlands[6,7]

1   Department of Primary Care & Public Health Sciences, School of Medicine, King's College London, UK
2   Department of Public Health, Section for Health Promotion and Health Services, Aarhus University, Denmark
3   Division of Women's Health, Faculty of Life Sciences & Medicine, King's College London, UK
4   Research Institute for Primary Care & Health Sciences, Keele University, UK
5   Center for Healthcare Studies, Institute for Public Health and Medicine and Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago IL, USA
6   Institute of Public Health, Aarhus University, Denmark
7   Institute of Health and Society, Newcastle University, UK

Corresponding Author: Karin R Laursen

✉   karinrosenkilde@gmail.com

Department of Primary Care & Public Health Sciences, School of Medicine, King's College London, UK

Tel: +45 22 32 22 68

Abstract

Context: Low health literacy (HL) is a public health issue, with impacts on population health and illness, however there are few tools for collecting health literacy data in large populations.

Objective: To develop a method of deriving indicative functional HL levels from routinely collected socio-demographic data.

Method: We investigated which socio-demographic variables would best depict whether an individual is above or below a constructed HL competency threshold. Weighted logistic regression was used to estimate Odd Ratios for being below the threshold. Weighted Receiver Operating Characteristic (ROC) analysis examined which variables best predicted low HL. Specificity, sensitivity and area under (AU) the ROC were descriptors for ability to predict risk.

Results: Three models were developed; one using all nine variables; a pragmatic model using the four most predictive variables (Qualification (whether the individual had achieved the level expected by age 16 years), Ethnicity, Home ownership, and Area Deprivation); and one using only "Qualification" (the single most predictive variable). All models showed good prediction of low HL (AUROC 0.73 (95% CI 0.71; 0.74) to 0.78 (95% CI 0.76; 0.79)), with predictive power increasing with more complex models.

Conclusion: The most important predictor of low HL is achievement of the qualification level expected by age 16 years, with additional variables adding more predictive power. The developed formulae can be used to estimate functional HL levels in populations from routinely collected socio-demographic data, and hence facilitate effective development and targeting of public health communications. The method to derive the formulae will be applicable in other industrialized countries.

Keywords: Health literacy; Epidemiology; Public health; Population characteristics

Introduction

The relationship between poor education, health literacy (HL) skills and health is well recognized [1-3]. Low health literacy is associated with greater use of medical services, lower use of preventative care, greater difficulty managing long term illnesses [1], lower levels of health [1-3] and higher mortality in older people [1, 2]. Further, it has been shown that public health messages fail to impact on those with low education, who are not only more likely to have fewer health-promoting behaviours, but are also less likely to respond to public health campaigns than their peers with higher education [4]. Public health campaigns may therefore

have the unexpected, and unwanted, effect of widening health inequalities [4]. Patient literacy and health literacy skills are thus of concern to those involved in communicating with patients and the public through public health campaigns to promote health and to reduce the risk of long-term and infectious disease, and in clinical settings to prevent and manage illness.

Low health literacy, "the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand, and use information in ways which promote and maintain good health" [5], is a social determinant of health. It is an independent predictor for poor health and mortality, augmenting the adverse effects of other social determinants of health such as membership of a minority ethnic group, poverty, limited education, and social deprivation [1-3, 6].

Increasing complexity of health care systems places greater cognitive demands on patients than ever before. Understanding complex health issues and obtaining necessary acute and preventive care requires understanding often complex information and navigating a system that requires a high level of understanding [7]. Materials provided to help people achieve and maintain health and to manage illness are too complex for literacy and numeracy skills of most people who need them [8].

Understanding the extent of the problems brought through low health literacy requires knowledge of the extent of the problem; in particular the number of people affected. Measurement of health literacy level at the individual level is one option. There are several validated measures of functional health literacy, capturing a wide range of health literacy skills. An example is the S-TOFHLA, a 12-minute test of literacy and numeracy skills that measures both health literacy and health numeracy skills levels, and enables people to be classified into 'inadequate', 'marginal' and 'adequate' health literacy levels [9]. There are, however, potential issues with direct measurement of HL levels, particularly individuals' possible feelings of stigma and inadequacy [10] and the time taken to complete measurement tests [11]. Another issue is that current measures are designed for individual, rather than community-level assessments, and provide little information about the level of health literacy within a population, unless applied in large population surveys. A model that could use currently available socio-demographic data to predict likely health literacy levels in individuals or populations would thus be useful for targeting clearer and more effective population public health communications.

Multiple reports have found high correlations between health literacy measures and demographic indicators such as age, ethnicity, and years of schooling [1, 3]. Imputed measures based on combinations of these indicators have been proposed [12, 13]. Miller et al. showed that an imputed measure correlated strongly with other indicators of health literacy among the elderly [12]. Hanchate et al. have developed and examined the performance of an imputed measure (the DAHL) of inadequate health literacy among elderly subjects for test-based measures commonly used in the literature [13]. Hanchate et al. aimed to give researchers a tool to explore the importance of health literacy in datasets where data had not been collected i.e. to develop a proxy for test-based measures. They showed that the DAHL captures most of those individuals who would be classified by the S-TOFHLA as having inadequate literacy [13].

The objective of the present study was built on the work of Hanchate et al. to develop and validate a method of deriving indicative functional health literacy levels from routinely collected socio-demographic data in a younger (working-age) population, applicable at national and international levels.

Method

Data

The following data sources were used: The '2011 Skills for Life Survey' [14] provided the socio-demographic variables whilst 'A mismatch between population health literacy and the complexity of health information: an observational study' [8] provided the health literacy and the combined health literacy and numeracy competency thresholds. A third data set, the '2003 Skills for Life Survey' [15] was used to validate the final models. This validation dataset was chosen because it contained exactly the same questions, assessments, recruitment and sampling strategy as the '2011 Skills for Life Survey', but a different population (**Table 1**).

2011 Skills for Life Survey, (SfL2011) SfL2011 was designed to measure basic skills amongst people aged between 16 and 65 in England. This was achieved by administering computerised assessments in literacy, numeracy and ICT (information and communication technology) to respondents during interviews. In all 7,230 interviews were conducted between May 2010 and February 2011. 6,050 individuals were assessed for literacy levels, 4,871 were also assessed for numeracy levels; whilst 2,274 were assessed for ICT levels. The population in SfL2011 consisted of even proportions of men and women (50% each), the majority of who categorised themselves as White British (80%). English was the first language for 89% of 16-65 year-olds. The population was distributed in roughly equal proportions across ten-year age bands [14].

A mismatch between population health literacy and the complexity of health information: an observational study

Rowlands et al. describe a method of measuring the gap between the complexity of health materials and the skills of the people for whom it is designed through identification of competency thresholds [8]. They identified two thresholds, one for health materials containing just text information, and one for health information containing both text (literacy) and numeracy information.

2003 skills for life survey, (SfL2003)

In these survey 8,730 randomly selected adults aged 16-56 were interviewed. In total 7,973 respondents completed the literacy test and 8,041 respondents completed the numeracy test. 7,517 completed both [15].

Table 1 Variables and data sources used in the present study.

Variables		Data sources		
Variable	**Description**	**SfL2011**	**SfL2003**	**Rowlands et al [8]**
Sex	Male, Female	✓	✓	
Age	16-65 years	✓	✓	
Ethnicity	White vs. Black and Ethnic Minority	✓	✓	
Language	Whether the first language is English	✓	✓	
Qualification level	English National Qualifications Framework (NQF) [20].	✓	✓	
Job status	National Statistics Socio-economic Classification	✓	✓	
Gross income (GBP)	Total personal income before tax and other deductions	✓	✓	
Home ownership	Whether a person owns or does not own his/her home	✓	✓	
Area deprivation	Index of multiple deprivation score [18]	✓	✓	
Health literacy threshold	The skills level required to understand and use 70% of health materials in common circulation.			✓
Health literacy & numeracy threshold				✓

The present study

Using the socio-demographic data from SfL2011 and the identified competency thresholds from Rowlands et al., [8] we investigated which variables would best depict whether an individual is above or below the competency thresholds.

Individuals in SfL2011, who has undertaken the literacy test, or the literacy and numeracy tests combined, were included in the study. Those individuals that did not start or finish the test were excluded leaving the final analytical samples of 5,824 individuals with established literacy levels of whom 4,773 also had an established numeracy level.

Ethics

All data used in this study are publically available and fully anonymised, and therefore ethics approval was not required. Confirmation was obtained from the Research Ethics Office at King's College London.

Statistical analysis

The two functional health literacy competency thresholds identified by Rowlands et al. (literacy and numeracy, and literacy (text) only) were the outcome variables for this study. Separate analyses were undertaken for these two outcomes. Descriptive characteristics were calculated for baseline demographics. The nine variables (sex, age, ethnicity, language, qualification level, job status, gross income, home ownership and area deprivation score) found to be statistically significant in Rowlands et al. were included in this study [8, 16].

In the SfL2011 and SfL2003 surveys, data were weighted to correct for sampling errors. Weighting was undertaken by comparing the socio-demographic profile of those allocated to receive the tests, and correcting for national profiles of the English working-age population. We used these weightings in our analysis. Weighted logistic regression was used to estimate odds ratios and z-values for being below the competency thresholds. Weighted univariable and multivariable Receiver Operating Characteristics (ROC) analysis combined with Bootstrap estimation was used to examine which variables had the largest area under the ROC curve. Bootstrap estimation was used to correct for the weights to ensure correct standard errors and confidence intervals.

Three models were developed for each outcome: one using all nine variables, a more pragmatic 4-factor model using only those variables in the logistic regression with high z-values and which are commonly collected in public surveys, and one using only the variable with the highest z-value and the largest area under the ROC curve. Models with interactions among the various variables were not explored. Specificity, sensitivity, likelihood ratios and area under the ROC were used as descriptors of each model's ability to predict an individual's risk of being below the competency threshold.

Validation

Two validation methods were used: a within-data (internal) validation and a between-data (external) validation. Initial within-data validation ensured our calibration was computed correctly and was applied to check that the prediction scores fitted the dataset. Subsequently, the models were validated against the SfL2003 dataset in order to get an unbiased assessment of how the models might perform in practice, as estimates from original datasets are typically over-optimistic [17].

To maximize methodological fidelity, the SfL2003 dataset was not downloaded until after the initial analysis generating models from the SfL2011 data. Variables were recoded as necessary to ensure they were equivalently specified in both the SfL2011 and the SfL2003 dataset. The validation was done using the same statistics as for the SfL2011 dataset. This was repeated for both literacy and numeracy tested and literacy-only tested individuals.

All the statistical analyses were performed using Stata Version 12. As this was an observational study, STROBE guidelines [18] were followed.

Results

Demographic characteristics for the combined literacy and numeracy-, and the literacy-tested, individuals are described in **Table 2**. Of the literacy and numeracy-tested individuals, 2,922/4,773 (61.2%) were below the competency threshold; of the literacy-tested individuals, 2,508/5,824 (43.1%) were below the competency threshold.

As most health materials contain both literacy and numeracy information [8], the results relating to the combined literacy and numeracy threshold are reported as the main findings, whilst the results of the literacy only threshold are reported in supplementary tables and figures.

The variable "qualification level" was found to be the single most predictive variable according to the z-value and the univariable ROC analysis. The additional three variables included in the pragmatic 4-factor model were ethnicity, home ownership, and socio-economic deprivation level of residential areas as measured by the Index of Multiple Deprivation (IMD) [19].

9 predictors	AUC=0.78 (95% CI0.767-0.793)
4 best predictors	AUC=0.77 (95% CI0.753-0.780)
Qualifications	AUC=0.73 (95% CI0.712-0.741)
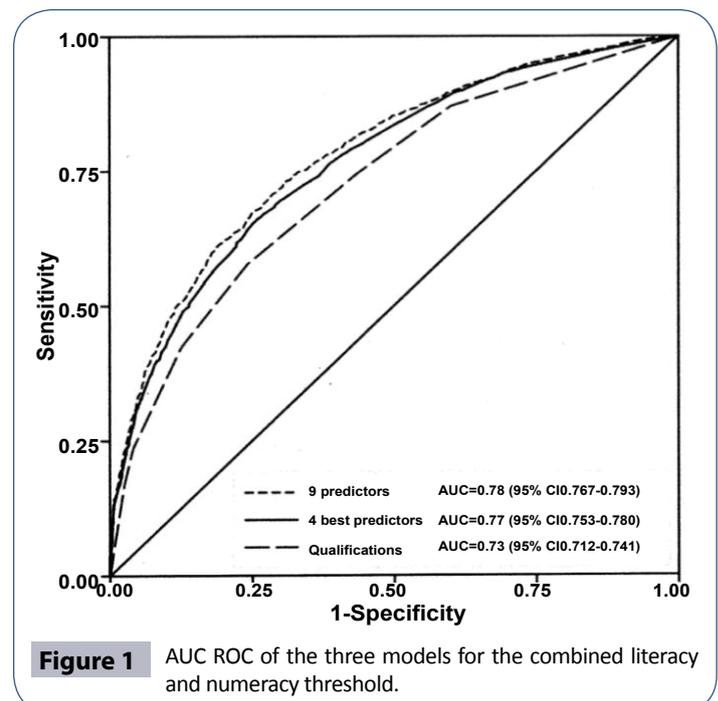
Figure 1 AUC ROC of the three models for the combined literacy and numeracy threshold.

Table 2 Demographic characteristics for participants in SfL2011.

	Literacy only tested	Literacy and numeracy tested
Tested individuals (n)	5,824	4,773
	Not literacy competent, n (%)	Not literacy+numeracy competent, n (%)
All	2,508 (43.1)	2,922 (61.2)
Gender: Male	1,132 (44.9)	1,205 (57.6)
Female	1,376 (41.7)	1,717 (64.1)
Missing	0 (0.0)	0 (0.0)
Age band: 16-24	322 (43.9)	407 (67.2)
25-34	460 (41.2)	547 (59.8)
35-44	481 (36.8)	591 (55.2)
45-54	583 (45.6)	640 (61.8)
55-65	660 (47.5)	735 (64.3)
Missing	2 (66.7)	2 (66.7)
Ethnic group: White	2,125 (40.9)	2,513 (59.1)
Black/Minority	382 (60.8)	407 (79.2)
Missing	1 (33.3)	2 (50.0)
Whether English is first language:		
Yes	2,190 (40.9)	2,607 (59.7)
No	318 (66.4)	315 (77.4)
Missing	0 (0.0)	0 (0.0)
Qualifications: Degree level	263 (19.3)	378 (33.8)
Non-degree level HE qualification	296 (35.2)	360 (54.5)
Level 3 qualification	379 (37.1)	498 (57.9)
Level 2 qualification	377 (46.9)	449 (67.5)
Level 1 qualification or below	471 (55.9)	550 (77.8)
Other qualification: level unknown	197 (71.4)	193 (87.3)
No qualification	525 (77.6)	494 (91.3)
Missing	0 (0.0)	0 (0.0)
NSSEC*: Managerial and professional	589 (26.9)	769 (43.5)
Intermediate occupation	454 (42.1)	570 (62.4)
Routine/manual, students and unemployed	1,465 (57.4)	1,583 (75.6)
Missing	0 (0.0)	0 (0.0)
Gross income less than 10,000 GBP:		
No	743 (32.3)	910 (48.4)
Yes	1,765 (50.1)	2,012 (69.5)
Missing	0 (0.0)	0 (0.0)
Owns or part-owns home: Yes	1,294 (35.5)	1,574 (52.9)
No	1,214 (55.7)	1,348 (75.1)
Missing	0 (0.0)	0 (0.0)
IMD score (Area deprivation): 0-9	354 (28.7)	465 (46.3)
10-19	726 (38.3)	863 (55.5)
20-29	463 (44.6)	553 (64.6)
30-39	395 (54.0)	428 (71.5)
40+	570 (61.8)	613 (80.9)
Missing	0 (0.0)	0 (0.0)

*NSSEC: UK Office for National Statistics Socio-economic Classification.

From **Figure 1**, it can be seen that Model 1 (9 predictors) had an AUC ROC of 0.78 (95% CI 0.767; 0.793), Model 2 (4 best predictors) had an AUC ROC of 0.77 (95% CI 0.753; 0.780), and Model 3 (Qualifications only) had an AUC ROC of 0.73 (95% CI 0.712; 0.741).

The 9-factor and the pragmatic 4-factor model were thus both significantly better than the one-factor model, while the 9-factor model was not significantly better than the 4-factor model.

Developed from 4,773 individuals with complete data, the formulae for each model predict an individual's log odds of being below the health literacy and numeracy threshold. The formula for being below the threshold is a function of a given persons characteristics (**Table 3**).

Table 4 shows each of the three model's diagnostic properties. They are based on a threshold that categorizes 61% of the population as having inadequate literacy and numeracy skills [8].

Table 3 Formulae for each of the three models[1].

	Model 1	Model 2	Model 3
Constant	-3.1880	-1.1961	-0.6357
Sex: Male Female	(ref.) 0.2899		
Age: 16-24 25-34 35-44 45-54 55-65	(ref.) 0.0970 0.1857 0.4078 0.2522		
Ethnicity: White Black/minority	(ref.) 0.7326	(ref.) 0.9906	
Language: English first language English not first language	(ref.) 0.6540		
Highest qualification: Degree level Non-degree level higher qualification Level 3 qualification (University entry) Level 2 qualification (expected level school leaving age) Level 1 qualification or below Other qualification No qualification	(ref.) 0.7624 0.7994 1.2037 1.7198 2.2300 2.5168	(ref.) 0.8877 0.9005 1.4061 1.9511 2.5931 2.8664	(ref.) 0.8472 0.8715 1.3949 1.9099 2.6688 2.9798
NSSEC: Managerial and professional Intermediate occupation Routine/manual, students and unemployed	(ref.) 0.2107 0.5073		
Gross income less than 10,000 GBP: No Yes	(ref.) 0.3220		
Owns or part-owns home: Yes No	(ref.) 0.4312	(ref.) 0.4808	
IMD (Area deprivation): 0-9 (least deprived) 10-19 20-29 30-39 40+	(ref.) 0.0559 0.3445 0.4585 0.7893	(ref.) 0.0698 0.3500 0.5182 0.8327	

[1]The predicted percentage probability of any participant in a study being below the health literacy and numeracy threshold can be calculated as: $e^L/(1+e^L)*100\%$. The formula for being below the threshold is thus Logit (pi) = (f(x))

Table 4 Prediction of low health literacy and numeracy[1].

	Model 1 (9 predictors)	Model 2 (4 predictors)[2]	Model 3 (Qualifications)
Sensitivity, %	76.6 (75.0;78.1)	75.0 (73.4;76.6)	57.7 (55.9;59.5)
Specificity, %	63.7 (61.5;65.9)	62.1 (59.9;64.3)	75.8 (73.8;77.7)
Positive predictive value, %	76.9 (75.3;78.4)	75.8 (74.2;77.3)	79.0 (77.2;80.7)
Negative predictive value, %	63.3 (61.1;65.5)	61.2 (58.9;63.4)	53.2 (51.2;55.1)
Likelihood ratio (+)	2.11 (1.98:2.25)	1.98 (1.86;2.11)	2.38 (2.19;2.60)
Likelihood ratio (-)	0.37 (0.34;0.40)	0.40 (0.37;0.43)	0.56 (0.53;0.59)
ROC curve area	0.70 (0.69;0.71)	0.69 (0.67;0.70)	0.67 (0.65;0.68)
Odds ratio (LR(+)/LR(-))	5.75 (5.06;6.53)	4.93 (4.34;5.59)	4.27 (3.75;4.86)

[1]All the numbers are stated as estimates with 95% confidence intervals.

[2]The four best health literacy and numeracy predictors: Qualifications, ethnicity, home ownership, and area deprivation level (IMD level).

Validation of the models

The internal validation proved that the prediction scores were calibrated correctly, and that the event rates for all the groups fell into the right categories. There was agreement between the probabilities and the observed data.

The results of the external validation using the SfL2003 dataset can be found in supplementary **Table 1S**. As expected, the predictions were less accurate compared to the original SfL2011 dataset, but are acceptable. The results suggest that, where data are available, the full 9-factor model should be used as it performed significantly better on the external validation dataset than the two other models (AUC ROC 0.73 (95% CI 0.717; 0.736) vs. respectively 0.70 (95% CI 0.686; 0.706) and 0.65 (95% CI 0.636; 0.657)).

For the literacy-only dataset, the outcomes of the modeling process were similar to the combined literacy and numeracy dataset. The factors in the 4-factor model, i.e., those with

6

the best predictive values for literacy-only competency were qualifications, age, language, and IMD. The 9-factor model had a ROC area of 0.76 (95% CI 0.747; 0.772), the 4-factor model had a ROC area of 0.75 (95% CI 0.733; 0.758), and the model with qualifications alone had a ROC area of 0.71 (95% CI 0.699; 0.725) (**Figure 1S**). The diagnostic properties of each model were tested and can be seen from **Table 2S and 3S**.

Discussion

Summary of findings

To our knowledge, this is the first model to predict health literacy in an English working-age population. Our method builds on that reported by Hanchate et al., which was developed and applied in a population of older people using US socio-demographic data [13].

For both the literacy and numeracy competency threshold and the literacy-only competency threshold, three models of varying complexity were developed to impute functional health literacy and numeracy levels. ROC areas between 0.71 and 0.78 indicated that, overall, the models discriminated well among people below and above the competency thresholds. The pragmatic 4-factor model was significantly better than the one-factor model, while the 9-factor model, whilst not significantly better than the 4-factor model in the SfL2011 dataset, did appear to perform better when tested against the external validation (SfL 2003) dataset. Each model and its related formula demonstrated fair to good diagnostic properties. The implications of these results are that education level, by far the strongest predictor of health-literacy competency, is essential in predicting health literacy levels; if education level is the only predictor available it will give a reasonable level of accuracy. If the additional three variables in the pragmatic model are available the accuracy of the prediction will be significantly improved. The improvement of the model still further, in the external validation dataset, by the 9-factor model, suggests that this is the best model to use if data on all the variables are available.

Strengths of the study

Strength of the study is the high quality of data. The dataset from which the models were developed (the SfL2011 and the validation dataset, SfL2003) are large nationally-representative samples of the English working-age population, using detailed individual-level socio-demographic data and literacy and numeracy assessments developed by education-testing experts; response bias is thus unlikely.

The availability of combined literacy and numeracy data on a large proportion of the survey sample enables models to be developed on combined literacy and numeracy skills. This is important as literacy and numeracy skills are not highly correlated at the individual level [14, 15] and most health materials contain both text (literacy) and numeracy information [8].

Limitations of the study

The models are limited by the explanatory power of respectively 9, 4 or 1 predictor variable(s) considered for respectively combined health literacy and numeracy, and health literacy alone. Although there are other cultural, societal, educational and health system factors that may improve the prediction of a person being below the threshold, this research was limited to explanatory variables commonly collected in public data sets.

In the Skills for Life surveys, population skills were measured using tests of a type that, whilst widely used in national and international surveys, have been criticised for only partially measuring skills, not adequately reflecting different cultures, and not adequately reflecting 'real life' [20]. However, the skills tests used have been extensively tested and validated, and provide the best available estimates of population literacy, numeracy, and health literacy skills in England.

Comparison with the existing literature

The DAHL is an imputed measure for community-living elderly aged 65 or older in USA [13]. Hanchate et al. based their models on the variables sex, age, ethnicity and years of schooling, which is similar to the socio-demographic indicators found to be predictive in our study. In our study, lower educational level, older age, and membership of a Black and Minority Ethnic Group (BME) resulted in lower imputed health literacy. Both the DAHL and the models developed in this study had similar AUC of approximately 0.8, indicating a good predictive power of the models.

Miller et al. [12] identified a model to predict self-reported health literacy levels in adults aged 65 and over. The predictive model used data on age, ethnicity, education and sex. This model correctly classified the health literacy level of 73% of their study sample. This compares to our 9-predictor and 4-predictor models classifying respectively 76.6% and 75.0% of the study sample correctly. A limitation of Miller's model is that all the health literacy data are derived from self-report, and therefore are potentially subject to response bias.

Implications for research and practice

The prediction formulae developed in this study enable a reasonably accurate prediction of health literacy competency from routinely collected socio-demographic data in the English working-age population. This enables researchers working on English datasets, where some or all of the variables in our model are collected, to derive indicative health literacy levels to population datasets, provided that the datasets contain data on educational level. Application of the formulae described in this paper to such datasets will enable researchers to explore the relationships between health literacy and health, education, and other social determinants of health. Such studies could include investigations of the health economic implications of health literacy, an area identified as an important area for research and development [1].

The models described in this paper could also be used to aid health service planning, particularly in developing clearer and more effective communication with patients and the public. Application of the models at borough area-level (150,000–350,000 people) or at national level could aid in identification of areas where services should be tailored for people with low health literacy skills, with redistribution of resources to enable health authorities in areas

with high numbers of people facing health literacy challenges to develop more effective services for their patients. Public health campaigns in these areas would require better tailoring of health promotion and disease prevention campaigns to improve impact, [4] with public health and health education practitioners and organisations trained to improve communication skills.

The similarity of our findings to those of Hanchate et al., [13] undertaken in the US, indicates that the method described here is likely to be applicable in most industrialised countries. Whilst the exact data collected, and the categorisation of those data, will vary between countries, it would appear that education level, age, whether the national language is an individuals' first language, and area socio-economic deprivation will be important and valid in national models.

Unanswered questions and future research

Future research should further explore ways to effectively improve communication with patients and the public, particularly those with lower health literacy skills, and evaluate the impact on patient satisfaction, patient safety, patient health, and the impact of public health campaigns. This study only addresses functional health literacy skills; studies that explore other health literacy skills, e.g., verbal communication skills, interactive and critical health literacy skills [11] would be very valuable.

Conclusion

This research has developed and validated a method for predicting population health literacy (and numeracy) levels from routinely collected socio-demographic data. The prediction models and formulas described in this paper can be used to further investigate health literacy, health and illness, and to manage health services to provide better health services to communicate better with people with low health literacy.

Acknowledgement

Key points

- Low health literacy has significant impact on public health but there are few tools for collecting health literacy data in large populations.

- As a result of this study, health literacy can be imputed for epidemiological datasets for adults of working-age.

- The method can be replicated in any country that has undertaken a health literacy survey and has other datasets with socio-demographic data.

- This will facilitate further research into the impact of low health literacy on population health, illness, and health care costs.

- It will also enable those commissioning health services to tailor resources to take account of local health literacy needs of the population.

References

1   Berkman ND, Sheridan SL, Donahue KE, David J Halpern, Anthony Viera, et al. (2011) Health literacy interventions and outcomes: An updated systematic review. Rockville, MD: Agency for Healthcare Research and Quality Contract No: 11-E006.

2   Bostock S, Steptoe A (2012) Association between low functional health literacy and mortality in older adults: longitudinal cohort study. British Medical Journal 344: e1602.

3   HLS-EU Consortium: Comparative report of health literacy in eight EU member states (2012) The European Health Literacy Survey (HLS-EU).

4   Buck D, Frosini F (2012) Clustering of unhealthy behaviours over time. Implications for policy and practice London: King's Fund.

5   Nutbeam D (1999) Health Promotion Glossary. Health promotion international 13: 349-364.

6   Baker DW, Wolf MS, Feinglass J (2007) Health literacy and mortality among elderly persons. Archives of Internal Medicine 167: 1503-1509.

7   US Department of Health and Human Services. Office of Disease Prevention and Health Promotion. National Action Plan to Improve Health Literacy. Washington D.C: 2010.

8   Rowlands G, Protheroe J, Winkley J, Richardson M, Seed PT, et al. (2015) A mismatch between population health literacy and the complexity of health information: an observational study. Br J Gen Pract 65: 379-386.

9   The National Qualifications Framework (2006).

10 Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J (1999) Development of a brief test to measure functional health literacy. Patient education and counselling 38: 33-42.

11 Baker DW, Parker RM, Williams MV, Pitkin K, Parikh NS (1996) The health care experience of patients with low literacy. Arch Fam Med 5: 329-334.

12 Nutbeam D (2000) Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health promotion international 15: 259-267.

13 Miller MJ, Degenholtz HB, Gazmararian JA, Lin CJ, Ricci EM, et al. (2007) Identifying elderly at greatest risk of inadequate health literacy: a predictive model for population-health decision makers. Res Social Adm Pharm 3: 70-85.

14 Hanchate AD, Ash AS, Gazmararian JA, Wolf MS, Paasche-Orlow MK. (2008) The Demographic Assessment for Health Literacy (DAHL): a new tool for estimating associations between health literacy and outcomes in national surveys. J Gen Intern Med 23: 1561-1566.

15 The 2011 Skills for Life Survey: A Survey of Literacy, Numeracy and ICT Levels in England. London: Department for Business Innovation and Skills 2012 Contract No: BIS Research paper number 81.

16 Williams J, Clemens S, Oleinikova K, et al. (2003) The skills for life survey. A national needs and impact survey of literacy, numeracy and ICT skills. London: Department for Education and Skills, UK.

17 Altman DG, Royston P (2000) What do we mean by validating a prognostic model?. Stat Med 19: 453-473.

18 Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, et al. (2007) Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 335: 806-808.

19 The English Indices of Deprivation 2010 (2011) Department for Communities and Local Government (England).

20 Hamilton M, Barton D (2000) The International Adult Literacy Survey: What Does It Really Measure? International Review of Education 46: 377-389.