

Machine learning using longitudinal prescription and medical claims for the detection of non-alcoholic steatohepatitis (NASH)

Ozge Yasar,¹ Patrick Long ,² Brett Harder,² Hanna Marshall,² Sanjay Bhasin,² Suyin Lee,² Mark Delegee,³ Stephanie Roy,² Orla Doyle,¹ Nadea Leavitt,² John Rigg¹

To cite: Yasar O, Long P, Harder B, *et al*. Machine learning using longitudinal prescription and medical claims for the detection of non-alcoholic steatohepatitis (NASH). *BMJ Health Care Inform* 2022;**29**:e100510. doi:10.1136/bmjhci-2021-100510

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjhci-2021-100510>).

OY and PL contributed equally.

Received 01 November 2021
Accepted 13 March 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Real World Solutions, IQVIA, London, UK

²Real World Solutions, IQVIA, Plymouth Meeting, Pennsylvania, USA

³Therapeutic Center of Excellence, IQVIA, Durham, North Carolina, USA

Correspondence to

Dr Patrick Long;
Patrick.long@iqvia.com

ABSTRACT

Objectives To develop and evaluate machine learning models to detect patients with suspected undiagnosed non-alcoholic steatohepatitis (NASH) for diagnostic screening and clinical management.

Methods In this retrospective observational non-interventional study using administrative medical claims data from 1 463 089 patients, gradient-boosted decision trees were trained to detect patients with likely NASH from an at-risk patient population with a history of obesity, type 2 diabetes mellitus, metabolic disorder or non-alcoholic fatty liver (NAFL). Models were trained to detect likely NASH in all at-risk patients or in the subset without a prior NAFL diagnosis (at-risk non-NAFL patients). Models were trained and validated using retrospective medical claims data and assessed using area under precision recall curves and receiver operating characteristic curves (AUPRCs and AUROCs).

Results The 6-month incidences of NASH in claims data were 1 per 1437 at-risk patients and 1 per 2127 at-risk non-NAFL patients. The model trained to detect NASH in all at-risk patients had an AUPRC of 0.0107 (95% CI 0.0104 to 0.0110) and an AUROC of 0.84. At 10% recall, model precision was 4.3%, which is 60× above NASH incidence. The model trained to detect NASH in the non-NAFL cohort had an AUPRC of 0.0030 (95% CI 0.0029 to 0.0031) and an AUROC of 0.78. At 10% recall, model precision was 1%, which is 20× above NASH incidence.

Conclusion The low incidence of NASH in medical claims data corroborates the pattern of NASH underdiagnosis in clinical practice. Claims-based machine learning could facilitate the detection of patients with probable NASH for diagnostic testing and disease management.

BACKGROUND

Non-alcoholic fatty liver disease (NAFLD) is an umbrella term that describes two subtypes of liver disease: non-alcoholic fatty liver (NAFL) and non-alcoholic steatohepatitis (NASH).^{1 2} NAFL is characterised by fat accumulation (steatosis) in the liver without significant inflammation. NASH is a more severe form of NAFLD and is characterised

Summary

What is already known?

- Non-alcoholic steatohepatitis (NASH) is difficult to detect without an invasive liver biopsy and is underdiagnosed despite the risk of progression to cirrhosis.
- Machine learning (ML) models trained on real-world data have shown promise in detecting rare or underdiagnosed diseases such as NASH.

What does this paper add?

- This study extends the existing literature on ML applications in healthcare by incorporating high coverage medical claims data as a scalable strategy for detecting patients with likely NASH.

How this study might affect research, practice or policy?

- This study may increase awareness of NASH underdiagnosis and under-reporting in clinical practice.
- This study may serve as a basis for future research aimed at validating ML models to support NASH diagnosis in the clinical setting.

by steatosis with inflammation and fibrosis, which can progress to cirrhosis.¹ The prevalence of NAFLD in the USA is estimated to be 24%–26% of adults, of whom an estimated 20%–30% have NASH.³

The transition from simple hepatic steatosis to NASH is a crucial point in the development of severe liver disease, putting patients at higher risk for fibrosis and progression to chronic liver disease.⁴ Nevertheless, NASH is often underdiagnosed in clinical practice.^{5–7} This may be due to several factors. First, there is a lack of clear patient symptoms and reliable biomarkers to help identify NASH,^{8 9} and there are no universal routine screening standards.¹⁰ Second, liver biopsy is the gold standard for NASH diagnosis but is costly, invasive, complicated by sampling errors

and requires a specialist to perform.^{4 11} Finally, despite ongoing clinical trials, there are currently no approved pharmacological treatments for NASH outside of India.¹² Thus, detection of NASH remains a challenge and reliable diagnostic tools, including minimal or even non-invasive techniques, are warranted.

Machine learning (ML) with real-world data may help address the underdiagnosis of common and rare diseases. We recently demonstrated the application of ML in a retrospective case-control cohort study based on a US claims database to identify patients with undiagnosed hepatitis C virus.¹³ For the detection of NASH, studies have yielded encouraging results using metabolomics,¹⁴⁻¹⁷ electronic health records¹⁸⁻²⁰ or combined clinical-claims data.²¹ The use case of each approach may be influenced by the chosen data type, characteristics of the model training population or the targeted application to patients with documented NAFL. Continuing to build on these efforts will further enable ML approaches to facilitate NASH detection.

This study examined supervised ML using medical claims as a non-invasive strategy to identify patients with likely NASH who might benefit from appropriate clinical follow-up such as monitoring or diagnostic screening. We used a retrospective rolling cross-sectional study design²² by taking multiple snapshots of patient prescription and medical claim histories to emulate patient data during real-world deployment while providing examples of patients with NASH for model training at different points in the patient journey prior to diagnosis. We also evaluated both knowledge-driven ('hypothesis-driven') and automated data-driven strategies in developing clinical predictors for NASH detection.

METHODS

Data sources

Data were extracted from IQVIA's proprietary US longitudinal prescription (LRx) and non-adjudicated medical claims (Dx) databases.¹³ LRx receives nearly 4 billion US prescription claims annually with coverage of 70%–90% of dispensed prescriptions from retail, mail order and long-term care channels. Dx receives over 1.35 billion US medical claims annually and covers approximately 70% of American Medical Association physicians. Dx data are derived from office-based individual professionals, ambulatory, general healthcare sites, hospitals, skilled nursing facilities and home health sites and includes patient-level diagnostic and procedural information.

All data were de-identified using an automated de-identification engine prior to being accessed by IQVIA. Both LRx and Dx data sets (hereafter referred to as LRxDx) are linked using anonymous patient tokens that support IQVIA data set interoperability and permits longitudinal linkage across patient histories. This anonymisation process is certified as Health Insurance Portability and Accountability Act compliant and Institutional Review Board exempt. IQVIA holds all requisite titles, licenses

and/or rights to license this Protected Health Information for use in accordance with applicable agreements. LRxDx data spanning the study period from 1 October 2015 to 30 June 2020 were used for cohort selection and predictive feature engineering. Dx data between 1 January 2010 and 1 October 2015 were used only to exclude patients with an existing International Classification of Disease-9 (ICD-9) NASH diagnosis before study initiation. This study was conducted using the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis reporting guidelines.

Patient selection

An overview of the patient populations used for cohort identification is shown in [figure 1A](#). Patients with a history of obesity, type 2 diabetes mellitus (T2DM), metabolic disorder or NAFL were selected to include individuals who also had precursors and risk factors for NASH.²³ This has the effect of enriching the patient population to ensure that an algorithm learns a prediagnosis footprint specific to NASH rather than simply distinguishing between healthy and sick patients or between patients with highly dissimilar symptomologies. This cohort was stratified to remove patients with claims for liver cancer, liver failure or alcohol-related liver disease and other liver complications that might disqualify patients from clinical intervention for NASH. Additional eligibility criteria were presence in both LRxDx for at least 24 months, recorded sex and age from 18 to 85 years at the time of model prediction. See online supplemental table 1 for cohort stratification criteria.

Rolling cross-section study design

To develop algorithms to detect NASH at different points of the prediagnosis patient journey (e.g., 1 month prediagnosis, 3 months prediagnosis), we divided patient claims history into a rolling series of time-bounded cross-sections (i.e., rolling cross-sections (RCS)). The application of this approach to disease detection is discussed in more detail elsewhere.²² Briefly, we divided longitudinal patient claims data over the study period into 24-month lookback periods followed by a 6-month outcome window, each shifted by 3-month increments ([figure 1B](#)). The lookback period was used to apply stratification criteria and extract data for feature engineering. The most recent date in the lookback period (the index date) represents the starting time point of model prediction.

Patients with probable NASH were labelled using the earliest occurrence of two criteria:

- ▶ A NASH diagnosis in the outcome window.
- ▶ A diagnosis for non-alcohol-related liver fibrosis, sclerosis or cirrhosis during the outcome window with a NAFL diagnosis within the preceding 24 or subsequent 6 months. These patients were presumed undiagnosed or unrecorded NASH patients as this clustering of diagnoses is indicative of NASH¹ and are therefore referred to as NASH proxy patients.

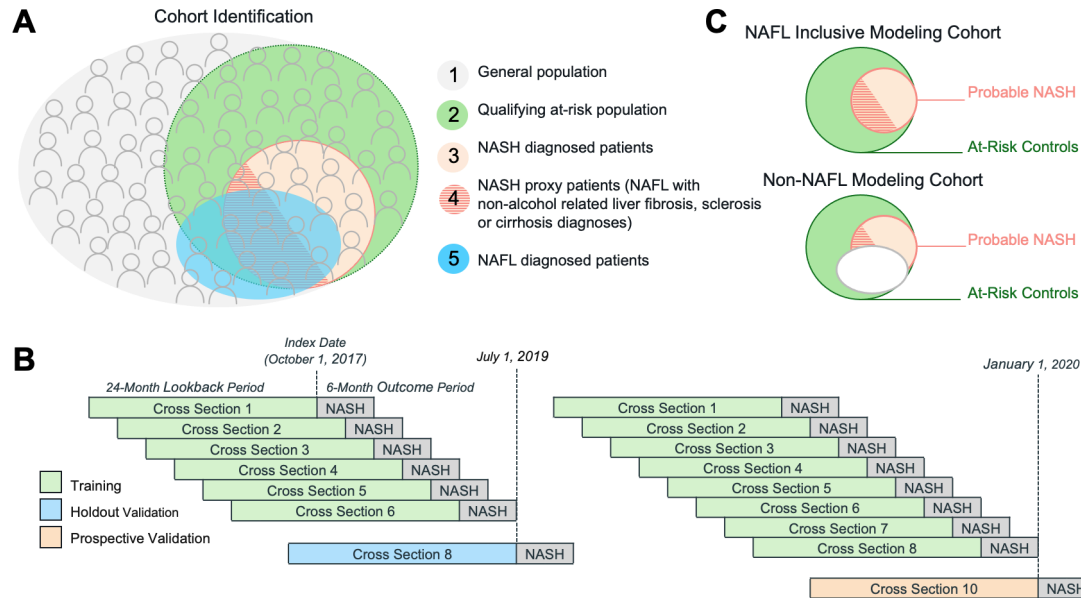


Figure 1 Cohort selection and model development. (A) Patients were identified from the general population (1) who met study stratification criteria (2). A subset of eligible patients was classified as likely to have NASH as evidenced by either a NASH diagnosis within 6 months after model prediction (3) or a diagnosis for non-alcohol-related liver fibrosis, sclerosis or cirrhosis within 6 months after model prediction and in proximity to a NAFL diagnosis (4). Eligible patients with no evidence of NASH were used as adversarial training examples, that is, patients with an overlap in symptoms, treatments and timing of resource utilisation but who were not patients with NASH. Patients with an existing NAFL claim in the previous 24 months overlapped with the patients with probable NASH population (5). Multiple time-bound cross-section were derived from the at-risk patient population (B). Cross-section 8 was reserved as a holdout set for model validation since it was sufficiently offset to prevent temporal overlap with the training set outcome window. Simulated model deployment using a prospective validation (scoring) set was performed by training on cross-sections 1–6 or on cross-sections 1–8 and using cross-section 10 as the test set. The NAFL inclusive and non-NAFL modelling cohorts used for model development (C). NAFL, non-alcoholic fatty liver; NASH, non-alcoholic steatohepatitis.

These patients had no other liver diagnosis for their cirrhosis.

See online supplemental table 2 for NASH labelling criteria.

A total of 152 476 patients met the selection criteria for patient with NASH labelling. Patients who met the selection criteria during the lookback period without evidence of NASH during or before the cross-section outcome window formed an at-risk control patient pool. A total of 54 976 837 at-risk control patients were initially identified and then randomly downsampled to a ratio of one patient with NASH per five at-risk control patients per cross-section to facilitate model training resulting in 1 312 351 at-risk patients for adversarial training.

Two modelling strategies were undertaken for NASH detection. The first (figure 1C, NAFL inclusive modelling) sought to detect NASH among all at-risk patients to maximise clinical impact. We also hypothesised that patients with documented NAFL might already be suspected of NASH and that this might limit the clinical utility of algorithm-based NASH screening. We therefore investigated a second modelling approach (figure 1C, non-NAFL modelling) by excluding patient cross-sections with a NAFL diagnosis claim during the lookback period.

Feature engineering

We used two methods for feature engineering. The first, a knowledge-driven (KD) approach, applied domain expertise to curate clinical codes into medical concepts associated with NASH risk such as relevant comorbidities, symptoms, procedures and treatments. The second, a data-driven (DD) approach, extracted clinical codes that were present in the NASH or at-risk control patient cross-section lookbacks. Codes of each type, for example, diagnoses or procedures, were then selected based on the largest absolute difference in prevalence between NASH and at-risk control patients with the motivation that these codes should be discriminatory predictors. DD feature identification was performed only with training cross-sections 1–6 to avoid data leakage from future cross-sections used for model validation. Patient demographics (age and sex) were included as model predictors to complement KD/DD feature sets and were included in all modelling scenarios.

Date differences and frequencies for claim and specialty visits were calculated over each patient lookback period. Date differences were calculated as the number of days between the first and last claim for a given feature relative to the index date and the number of days between the first and last claim. Missing data were represented as zero for frequency features and as null for date difference

features since null values are handled inherently by the ensemble algorithm used. See online supplemental tables 3-5 for clinical concepts used to derive model predictors.

Model selection

Gradient boosted trees²⁴ using the XGBoost package²⁵ were trained to discriminate between NASH and at-risk control patients. XGBoost was selected based on its previous success in benchmarking against other disease detection algorithms using claims data,^{13 26} its suitability over deep learning for tabular data²⁷ and its compatibility with sparse claims data sets and scalability.²⁵ Training cross-sections 1-6 were used for recursive feature elimination for feature selection (online supplemental figure 1) and for hyperparameter optimisation using grid search (online supplemental table 6).

Model evaluation

Models were evaluated using the area under the precision recall curve (AUPRC).²⁸ To compensate for random downsampling of the at-risk control cohort, precision was scaled to the 6-month incidence of NASH in the at-risk patient population before downsampling, which was 1 per 1437 and 1 per 2127 patients for the NAFL inclusive and non-NAFL cohorts, respectively. Such scaling ensures that the number of false positives used to calculate precision is not underestimated and accurately reflects the incidence of NASH captured in claims data. Model precision using 95% CIs were approximated by treating each recall decile as a binomial distribution with *n* patients and a Bernoulli trial success probability of *p*. We then determined the uncertainty of *p* using a beta distribution with true and false positive predictions. The receiver operating characteristic curve and the corresponding area under the curve (AUROC) are given for reference. Feature importance was examined using SHAP (SHapley Additive exPlanations).²⁹

Algorithms were compared with screenings for NASH using evidence of NAFL or T2DM in the last 2 years of a patient's claims history (see online supplemental table 7 for qualifying clinical codes). Precision and recall of NAFL or T2DM screening were measured in NAFL inclusive and non-NAFL holdout sets. Model precision was then measured at the corresponding recall of each screening. The non-NAFL model was compared only to screening with T2DM.

RESULTS

Study cohort

The NAFL inclusive and non-NAFL modelling cohorts displayed similar clinical profiles. T2DM was more common in patients with NASH (67.1% NASH and 56.3% at-risk and 70.4% NASH and 56.4% at-risk patients in the NAFL inclusive and non-NAFL cohorts, respectively). Obesity was common in patients with NASH and in at-risk controls regardless of NAFL history (59.8% vs 59.3%).

NAFL was 10-fold more common in patients with NASH compared with at-risk controls (table 1).

Model performance

NAFL inclusive modelling detected patients with probable NASH in the at-risk holdout population with 0.0107 AUPRC (95% CI 0.0104 to 0.0110) and 0.84 AUROC (figure 2A,B). At 10% recall, the model detected patients with probable NASH with 4.3% precision (figure 2A). This represents a 60-fold improvement over the 6-month incidence of NASH, that is, if at-risk patients were randomly screened for NASH (figure 3A). Patients labelled using a NASH diagnosis or NASH proxy criteria were detected with an AUPRC of 0.0059 and 0.0051, respectively (online supplemental figure 3).

NAFL is a precursor of NASH; however, only 35% of patients with NASH received a NAFL diagnosis claim during their 24-month lookback period (table 1). Therefore, we developed a second model to detect patients with NASH in the non-NAFL cohort. The non-NAFL model identified patients with probable NASH with 0.0030 AUPRC (95% CI 0.0029 to 0.0031) and 0.78 AUROC (figure 2C,D). At 10% recall, this model detected patients with probable NASH with 1% precision, a 20-fold improvement over the incidence of NASH in this cohort (figure 3B). A comparable number of patients with NASH without NAFL were detected at approximately 30% recall by the NAFL inclusive model.

For both NAFL inclusive and non-NAFL cohorts, model precision and recall in the holdout set (i.e., train on cross-sections 1-6, test on 8) were comparable to that of the prospective validation set (i.e., train on cross-sections 1-8, test on 10) (figure 2A). In addition, models trained on cross-sections 1-6 and then tested on either the holdout or prospective validation set showed comparable performance (online supplemental figure 2), indicating model stability over a 6-month period.

ML surpassed the precision of NASH detection when screening at-risk patients with NAFL or T2DM claims and is presented as a fold improvement over the baseline incidence of NASH in claims data. In the NAFL inclusive cohort, NAFL screening detected 36% of likely patients with NASH with 9.9-fold precision (figure 3C), whereas the NAFL inclusive model detected the same number of patients with NASH (i.e. equivalent recall) with 16.1-fold precision (figure 3C). Screening with T2DM detected 66% of patients with likely NASH with 1.2-fold precision versus 4.5-fold precision with the NAFL inclusive model (figure 3D). In the non-NAFL cohort, T2DM screening detected 70% of patients with likely NASH with 1.3-fold precision, whereas the non-NAFL model improved precision by 2.3-fold for the same recall (figure 3E).

To evaluate the relative effectiveness of each feature engineering strategy, we compared the performance of models optimised with a mixture of KD and DD features to those optimised with one of the two feature types. Models developed with DD features performed slightly better than models developed with KD features. However,

Table 1 Cohort statistics

	NAFL inclusive modelling		Non-NAFL modelling	
	NASH	At-risk controls	NASH	At-risk controls
Patients (n)	152 476	1 312 351	104 219	1 265 649
Patient cross-sections (n)	265 785	1 328 897	172 423	1 281 376
Demographics				
Age in years, mean (SD)	57.3 (13.4)	55.8 (15.7)	57.6 (13.4)	55.8 (15.9)
Sex (female) (%)	59.8	59.4	59.0	58.0
NASH identification criteria (%)				
NASH ICD-10	70.3	0	76.7	0
NASH proxy	29.7	0	23.3	0
Inclusion criteria (%)				
Type 2 diabetes	67.1	56.3	70.4	56.4
Obesity	58.9	59.3	58.9	59.3
Metabolic syndrome	2.5	1.3	2.4	1.3
NAFL	35.1	3.6	0	0
Comorbidities (%)				
Hypertension	64.0	49.4	62.6	49.0
Morbid (severe) obesity	17.4	10.0	17.0	9.7
Abnormal results of liver function studies	14.0	1.5	9.2	1.2
Abnormal levels of other serum enzymes	12.3	1.8	8.2	1.5
Procedures (%)				
Liver biopsy	1.0	<1.0	<1.0	<1.0
Liver panel	14.7	5.2	10.5	4.9
Abdominal ultrasound	29.1	5.7	14.6	4.1
Comprehensive metabolic panel	9.9	7.3	9.2	7.2
Specialty visits (%)				
Gastroenterology	30.9	12.7	22.4	12.0
Endocrinology	10.9	6.1	10.0	6.0
Cardiology	30.4	20.6	28.2	20.2

Diagnoses, procedures and physician specialty visit information is derived from medical and prescription claims data captured during each patient cross-section lookback period.

ICD-10, International Classification of Disease-10; NAFL, non-alcoholic fatty liver; NASH, non-alcoholic steatohepatitis.

detection of NASH was maximised using the combination of KD and DD features (figure 2).

Model interpretation

To gain insight into which clinical factors drive algorithmic detection of NASH, we examined feature importance using SHAP (figure 4). As attributes (e.g., claim frequency and claim recency) may be correlated within a single feature, we ranked the top features by taking the cumulative sum of the absolute SHAP values for attributes within each feature. For the NAFL inclusive model, a prior NAFL diagnosis was the dominant clinical predictor accounting for 13% of the total contribution (figure 4A). In contrast, the top features were more evenly distributed in the non-NAFL model (figure 4B). Although trained independently, KD and DD models relied on similar clinical event types including comorbidities (T2DM),

laboratory findings (abnormal liver function studies and abnormal serum enzyme levels) and diagnostic procedures (abdominal ultrasonography) (figure 4C–F).

DISCUSSION

This study provides encouraging results for the use of medical claims data and ML to detect patients with likely NASH from large at-risk patient populations. Although there are no universally accepted routine screening standards for NASH,¹⁰ both NAFL and T2DM are well-recognised risk factors.² Nonetheless, claims-based algorithms outperformed NASH screening using NAFL or T2DM history alone. In addition, algorithms detected probable NASH in at-risk patients without documented NAFL, potentially representing an overlooked NASH

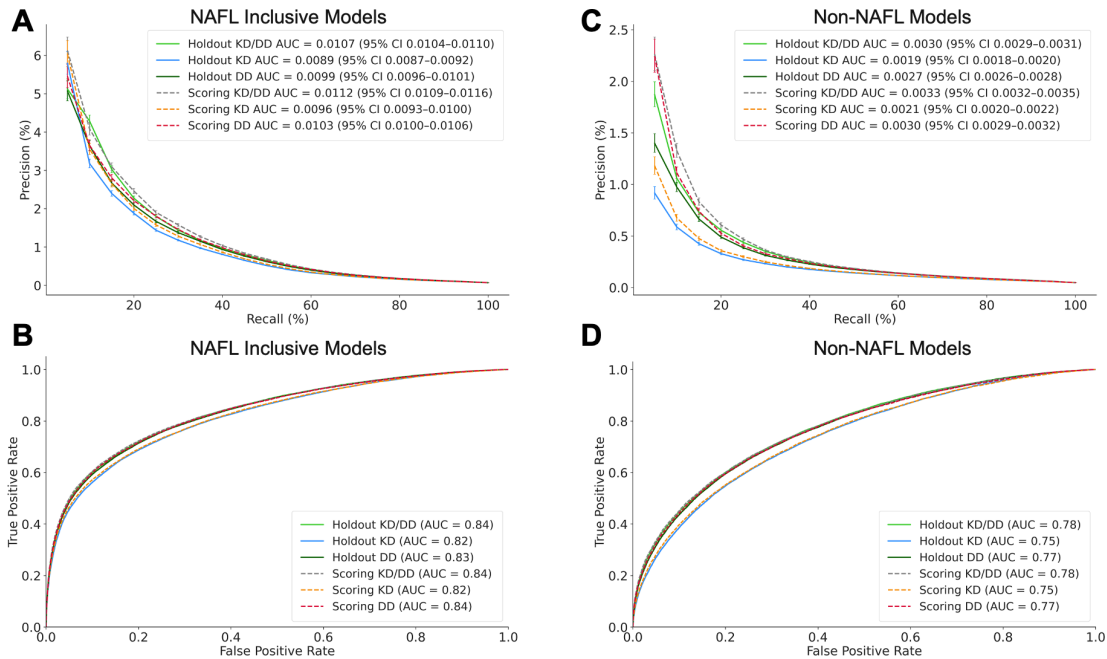


Figure 2 Model performance. Precision recall and receiver operating characteristic curves for model performance in detecting NASH in the NAFL inclusive (A and B) and non-NAFL (C and D) holdout and scoring validation sets. Precision is scaled to the 6-month NASH incidence observed in claims data. AUC, area under the curve; DD, data-driven; KD, knowledge-driven; NAFL, non-alcoholic fatty liver; NASH, non-alcoholic steatohepatitis.

risk group. Approaches such as this may support more targeted screening of prevalent and underdiagnosed diseases and may be particularly valuable when diagnosis

requires invasive or costly procedures or when clinical risk factors that could be used to screen patients are imprecise.

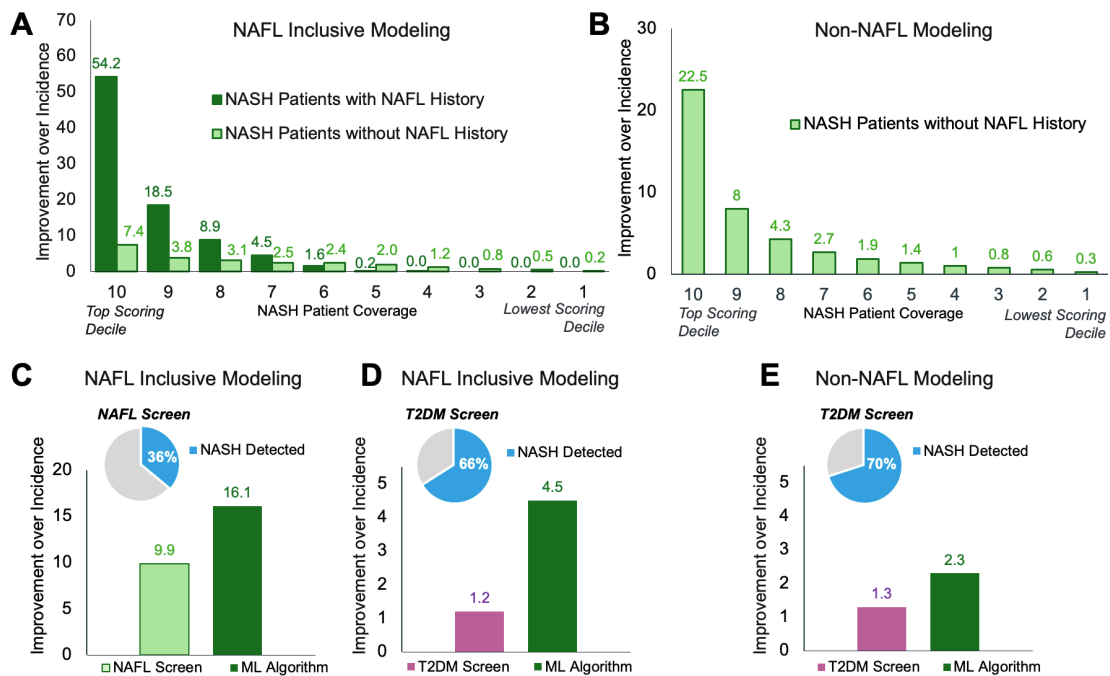


Figure 3 Model benchmarking. The fold improvement in model precision over the 6-month incidence of NASH observed in medical claims data and proportions within recall deciles of patients with predicted NASH with and without a NAFL diagnosis during the lookback period (A and B). Benchmark comparisons between NASH detection by ML algorithms and NASH screening using NAFL (C) or T2DM (D and E). The fold improvement over precision is calculated as the precision within each recall quantile divided by NASH incidence. ML, machine learning; NAFL, non-alcoholic fatty liver; NASH, non-alcoholic steatohepatitis; T2DM, type 2 diabetes mellitus.

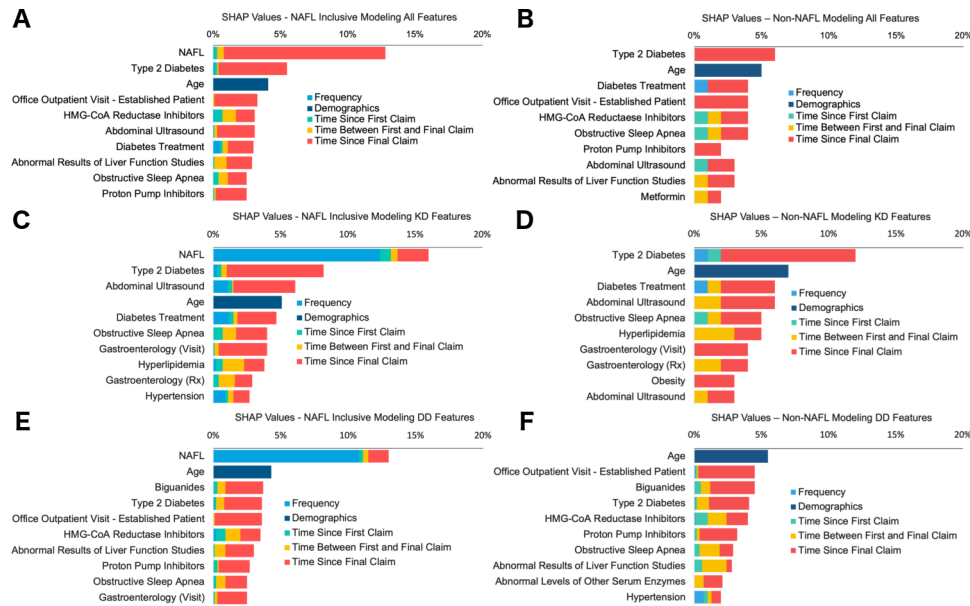


Figure 4 Feature Importance. SHAP values for top model features and the contribution of feature attributes, that is, claim frequency, time from first and final claim relative to the cross-section index date, time between the first and final claim occurrence and patient demographics for NAFL inclusive and non-NAFL combined KD/DD feature models (A–B), KD features models (C–D) and DD features models (E–F). SHAP values are expressed as a percentage of the total mean absolute SHAP values for models deployed on the holdout validation set. DD, data-driven; KD, knowledge-driven; NAFL, non-alcoholic fatty liver; NASH, non-alcoholic steatohepatitis; SHAP, SHapley Additive exPlanations.

We investigated two methods that may broadly inform ML healthcare applications. First, we evaluated an automated DD approach to feature engineering for algorithmic disease detection. Model performance was greatest when KD and DD features were combined, suggesting that the two feature engineering methods make complementary contributions by integrating clinical and epidemiological knowledge with an empirically oriented process of scientific inquiry. Such automated approaches may improve knowledge discovery in real-world data while reducing the burden on clinical domain experts. Second, our RCS method can facilitate the creation of cohorts using longitudinal health data and provide opportunities to monitor healthcare market dynamics that may impact model performance.

These models represent claims-based screening tools to facilitate the identification of patients with likely NASH who may benefit from clinical follow-up (e.g., via FibroScan) and as proof of concepts for further clinical validation. Potential users of these models include providers or payers who wish to implement high volume screening for suspected NASH in an at-risk patient population. While the NAFL inclusive model is suited for broad NASH detection, the non-NAFL model may be appropriate for screening patients for NASH for whom NAFL status is not well documented or a reliable cause of medical care.

There are several limitations in this study. First, model precision is likely underestimated due to NASH underdiagnosis and under-reporting,^{5–7} which may inflate the false positive rate in model evaluation. Changes in clinical practice that facilitate NASH diagnosis should close the gap between observed incidence in claims and epidemiological estimates while also enabling the development of

more powerful claims-based models. Second, our NASH labelling criteria do not guarantee NASH, which requires histological confirmation with a liver biopsy. The low percentage of patient cross-sections with a liver biopsy claim in this study may be due in part to limited coverage of this procedure in this data set. In addition, liver biopsy may not be performed in all cases, as a 2014 survey found that less than 25% of gastroenterologists and hepatologists routinely perform a biopsy to confirm NASH diagnosis.⁶ Clinical validation of these models would need to be performed on patients with liver biopsy-confirmed NASH and ideally assessed using multiple distinct medical claims data sets. Third, models were trained to detect patients with probable NASH regardless of NASH stage. Additional data types may enable stage specific NASH labelling for more targeted clinical interventions. Fourth, this study used a 6-month outcome window for NASH detection, which was chosen to allow indexing on more recent claims data. However, progressive diseases such as NASH may also benefit from longer prediction horizons to enable earlier detection. Finally, the robustness of our feature engineering strategy should be determined in subsequent sensitivity analyses using a broader range of techniques such as cost-sensitive or semi-supervised learning to address class imbalances³⁰ as well as alternative ML algorithms and clinical targets.

CONCLUSIONS

In this study, we investigated claims-based ML as a non-invasive and scalable approach to stratify patients with probable NASH from an at-risk population for clinical follow-up.

We also demonstrated automated DD feature engineering and an RCS study design in the development of disease detection algorithms. Models from this study or derivatives thereof could support more precise screening for NASH and help connect patients with available and emerging therapies.

Acknowledgements We are very grateful to Paranjoy Saharia, HEOR Scientific Services, IQVIA, and Rehan Ali, PhD and Benjamin North, PhD, Real World Solutions, IQVIA, for manuscript support.

Contributors Conception and design of the study and interpretation of the results: all authors. Data collection and analysis: OY, PL and BH. Original draft of the manuscript: PL. Critical revision and final review of the manuscript: all authors. PL acts as the guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests All authors are employees of IQVIA.

Patient consent for publication Not applicable.

Ethics approval The claims data used in this study were previously collected and statistically deidentified and are compliant with the deidentification conditions set forth in Sections 164.514 (a)-(b)1ii of the Health Insurance Portability and Accountability Act of 1996 Privacy Rule. No direct patient contact or primary collection of individual human patient data occurred. Study results were in tabular form and aggregate analyses, which omitted patient identification information. As such, the study did not require institutional review board review and approval or patient informed consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. All data belongs to IQVIA.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Patrick Long <http://orcid.org/0000-0002-7206-4607>

REFERENCES

- Chalasanani N, Younossi Z, Lavine JE, *et al*. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American association for the study of liver diseases. *Hepatology* 2018;67:328–57.
- Younossi ZM, Koenig AB, Abdelatif D, *et al*. Global epidemiology of nonalcoholic fatty liver disease—Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016;64:73–84.
- Shetty A, Syn W-K. Health and economic burden of nonalcoholic fatty liver disease in the United States and its impact on veterans. *Fed Pract* 2019;36:14–19.
- Drescher H, Weiskirchen S, Weiskirchen R. Current status in testing for nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH). *Cells* 2019;8:845.
- Alexander M, Loomis AK, Fairburn-Beech J, *et al*. Real-World data reveal a diagnostic gap in non-alcoholic fatty liver disease. *BMC Med* 2018;16:130.
- Rinella ME, Lominadze Z, Loomba R, *et al*. Practice patterns in NAFLD and NASH: real life differs from published guidelines. *Therap Adv Gastroenterol* 2016;9:4–12.
- Loomba R, Wong R, Frayssse J, *et al*. Nonalcoholic fatty liver disease progression rates to cirrhosis and progression of cirrhosis to decompensation and mortality: a real world analysis of Medicare data. *Aliment Pharmacol Ther* 2020;51:1149–59.
- Atabaki-Pasdar N, Ohlsson M, Viñuela A, *et al*. Predicting and elucidating the etiology of fatty liver disease: a machine learning modeling and validation study in the Iml direct cohorts. *PLoS Med* 2020;17:e1003149.
- Chan T-T, Wong VW-S. In search of new biomarkers for nonalcoholic fatty liver disease. *Clin Liver Dis* 2016;8:19–23.
- Pandeyarajan V, Gish RG, Alkhoury N, *et al*. Screening for nonalcoholic fatty liver disease in the primary care clinic. *Gastroenterol Hepatol* 2019;15:357–65.
- Nalbantoglu ILK, Brunt EM. Role of liver biopsy in nonalcoholic fatty liver disease. *World J Gastroenterol* 2014;20:9026–37.
- Sharma M, Premkumar M, Kulkarni AV, *et al*. Drugs for non-alcoholic steatohepatitis (NASH): quest for the Holy Grail. *J Clin Transl Hepatol* 2021;9:40–50.
- Doyle OM, Leavitt N, Rigg JA. Finding undiagnosed patients with hepatitis C infection: an application of artificial intelligence to patient claims data. *Sci Rep* 2020;10:10521.
- Canbay A, Kälsch J, Neumann U, *et al*. Non-Invasive assessment of NAFLD as systemic disease—A machine learning perspective. *PLoS One* 2019;14:e0214436.
- Perakakis N, Polyzos SA, Yazdani A, *et al*. Non-Invasive diagnosis of non-alcoholic steatohepatitis and fibrosis with the use of omics and supervised learning: a proof of concept study. *Metabolism* 2019;101:154005.
- Goldman O, Ben-Assuli O, Rogowski O, *et al*. Non-Alcoholic fatty liver and liver fibrosis predictive analytics: risk prediction and machine learning techniques for improved preventive medicine. *J Med Syst* 2021;45:22.
- Perveen S, Shahbaz M, Keshavjee K, *et al*. A systematic machine learning based approach for the diagnosis of non-alcoholic fatty liver disease risk and progression. *Sci Rep* 2018;8:2112.
- WY SPB, Xiao C, Glass L, *et al*. *Clifford g.d. a deep learning approach for classifying nonalcoholic steatohepatitis patients from nonalcoholic fatty liver disease patients using electronic medical records. Explainable AI in healthcare and medicine studies in computational intelligence*. Cham: Springer, 2021.
- Danford CJ, Lee JY, Strohbehn IA, *et al*. Development of an algorithm to identify cases of nonalcoholic steatohepatitis cirrhosis in the electronic health record. *Dig Dis Sci* 2021;66:1452–60.
- Docherty M, Regnier SA, Capkun G, *et al*. Development of a novel machine learning model to predict presence of nonalcoholic steatohepatitis. *J Am Med Inform Assoc* 2021;28:1235–41.
- Fialoke S, Malarstig A, Miller MR, *et al*. Application of machine learning methods to predict non-alcoholic steatohepatitis (NASH) in non-alcoholic fatty liver (NAFL) patients. *AMIA Annu Symp Proc* 2018;2018:430–9.
- Malpede B, Roy S, Long P, *et al*. AI plus real-world data for early prediction of disease progression and Operationalized precision targeting. *PMSA* 2020;8.
- Angulo P. Obesity and nonalcoholic fatty liver disease. *Nutr Rev* 2007;65:57–63.
- Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann. Statist.* 2001;29:1189–232.
- Association for Computing Machinery. *Xgboost: a scalable tree boosting system. knowledge discovery and data mining; 2016 August*. San Francisco, CA, USA, 2016.
- Morel D, Yu KC, Liu-Ferrara A, *et al*. Predicting Hospital readmission in patients with mental or substance use disorders: a machine learning approach. *Int J Med Inform* 2020;139:104136.
- Shwartz-Ziv R, Armon A, Ravid Shwartz-Ziv AA. Tabular data: deep learning is not all you need. *Information Fusion* 2022;81:84–90.
- Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015;10:e0118432.
- ACM Digital Library. A unified approach to interpreting model predictions. *31st conference on neural information processing systems (NIPS 2017)*, Long Beach, CA, USA, 2017.
- Krawczyk B. Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence* 2016;5:221–32.

© 2022 Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ. <http://creativecommons.org/licenses/by-nc/4.0/>This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>. Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.

'Refbin' an online platform to extract and classify large-scale information: a pilot study of COVID-19 related papers

Shania Lunna,¹ Isabelle Flinn,¹ James Prytherch,² Camille Torfs-Leibman,¹ Sarah Robtoy,¹ Matt Bansak,² David Krag ³

To cite: Lunna S, Flinn I, Prytherch J, *et al.* 'Refbin' an online platform to extract and classify large-scale information: a pilot study of COVID-19 related papers. *BMJ Health Care Inform* 2022;**29**:e100452. doi:10.1136/bmjhci-2021-100452

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjhci-2021-100452>).

Received 30 August 2021
Accepted 13 February 2022

ABSTRACT

Introduction The number of new biomedical manuscripts published on important topics exceeds the capacity of single persons to read. Integration of literature is an even more elusive task. This article describes a pilot study of a scalable online system to integrate data from 1000 articles on COVID-19.

Methods Articles were imported from PubMed using the query 'COVID-19'. The full text of articles reporting new data was obtained and the results extracted manually. An online software system was used to enter the results. Similar results were bundled using note fields in parent-child order. Each extracted result was linked to the source article. Each new data entry comprised at least four note fields: (1) result, (2) population or sample, (3) description of the result and (4) topic. Articles underwent iterative rounds of group review over remote sessions.

Results Screening 4126 COVID-19 articles resulted in a selection of 1000 publications presenting new data. The results were extracted and manually entered in note fields. Integration from multiple publications was achieved by sharing parent note fields by child entries. The total number of extracted primary results was 12 209. The mean number of results per article was 15.1 (SD 12.0). The average number of parent note fields for each result note field was 6.8 (SD 1.4). The total number of all note fields was 28 809. Without sharing of parent note fields, there would have been a total of 94 986 note fields.

Conclusion This pilot study demonstrates the feasibility of a scalable online system to extract results from 1000 manuscripts. Using four types of notes to describe each result provided standardisation of data entry and information integration. There was substantial reduction in complexity and reduction in total note fields by sharing of parent note fields. We conclude that this system provides a method to scale up extraction of information on very large topics.

INTRODUCTION

With >1 000 000 new articles per year^{1 2} it is nearly impossible to comprehensively assimilate newly published literature.^{3 4} PubMed and Google Scholar are powerful search tools, but the product is only a set of citations. Systematic reviews have a short life cycle⁵ and the process of establishing continuity between reviews is not well defined. Tools that facilitate reading,

describing and integrating data are limited. Linking results across multiple manuscripts is difficult.³ Tagging articles with keywords may help bundle sets of manuscripts but is a poor substitute for quality extraction of results.

We present pilot results extracting and integrating data from 1000 COVID-19 publications using a new online system and discuss the scalability and public dissemination of the information.

METHODS

A single account was used for the COVID-19 database (Refbin.com, Plomics, Shelburne, Vermont). Refbin retrieved citations from PubMed using COVID-19 as the search term. Articles that reported new data were selected (1000 of 4236). The results were manually extracted from full-text articles. Reviewers included six undergraduates, five postgraduate and three faculty. The average number of articles per reviewer was 132 (SD 243).

The extracted results were entered manually into the COVID-19 library as independent units of information. Each result was described by a set of note fields. The note fields were arranged in parent-child order and included (1) result, (2) description of the result, (3) population and (4) topic. The result note field occupied the lowest child note field. The citation was dragged to the result note field and was automatically linked to all parent note entries (see online supplemental methods for additional details).

RESULTS

From 1000 publications, 12 209 individual results were extracted. These were organised under 15 first-level topic note fields (online supplemental figure 1). The order, pattern and total number of parent notes varied for each extracted result (figure 1). The average number of note fields required to describe an



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹University of Vermont, Burlington, Vermont, USA

²Plomics, Shelburne, Vermont, USA

³Surgery, University of Vermont, Burlington, Vermont, USA

Correspondence to

Dr David Krag;
david.krag@uvm.edu

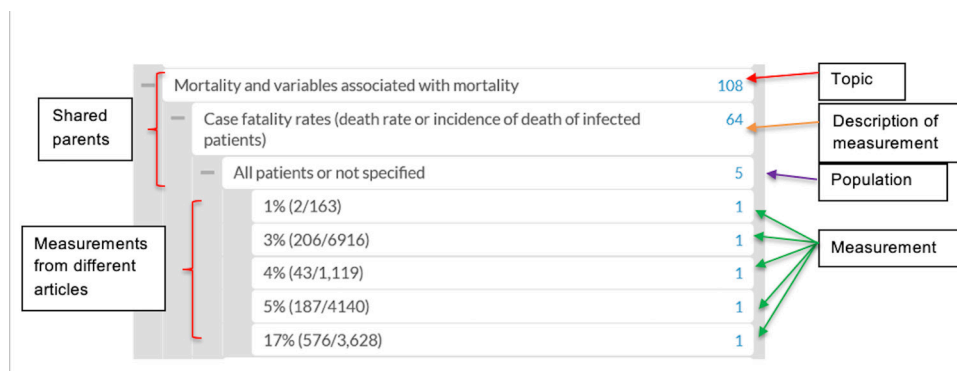


Figure 1 Screenshot of multiple entries of case fatality rates at the same hierarchical level sharing the same parents and four types of note fields. The number on the right edge of the note field is a citation counter. The parent note fields display the sum of citations affiliated with its children.

extracted result was 7.78 (SD 1.42) (online supplemental table 1).

The results from one manuscript were entered into multiple relevant locations throughout the database. The results from multiple manuscripts with similar features, for example case fatality rates, were bundled together. These similar results were entered into the same note field level and then could share parents (figure 1).

Sharing of parent note fields allowed substantial reduction in the total number of note fields and simplification of data entry. Without sharing of parents, the total number of all note fields would have been 94986. With sharing of parents, the total number of note fields was reduced ~70% to 28613 total note fields. This represents an absolute reduction of 66373 note fields (online supplemental figure 2).

To assess communication tools, 719 corresponding authors from 1000 publications were invited by email to obtain a personal Refbin account that included a read-only live portal to the COVID-19 library. This was accepted by 21% (148 of 719) and represented authors from 22 different countries. A publicly available website (COVIDpublications.org) was also created that displayed a live version of the COVID-19 library.

DISCUSSION

Four types of note fields were sufficient to describe each of the wide variety of results extracted from 1000 articles. Similar results shared parent note fields. This facilitated integration of data by creating frameworks to bundle results from different articles. This resulted in 70% reduction in the total number of note fields.

Online review methods minimised the frequency of scheduled meetings. Online group discussions were reserved for more challenging issues. Treating the extracted results as independent units of information allowed the results from one article to be placed in different library locations. This also allowed similar results from multiple articles to be integrated into a single location.

This pilot study describes a system that facilitates multiple individuals to read, describe and integrate results in a scalable manner. A live portal to the COVID-19 library was initiated for 148 researchers from 22 countries. A public version of the COVID-19 library was accomplished (COVIDpublications.org). This sets the stage for researchers to work together on areas of common interest and enhanced sharing of information across international borders.

Limitations of this study include dependency on the narrative language to describe a result. This limitation is mitigated by using the four note field types to describe each result, which helps standardise descriptions. Manual extraction of information is time-consuming. However, by providing the information as in this pilot study, time investment by a larger population of users may potentially be reduced. Reproducibility is a potential limitation. This was somewhat mitigated in this pilot study by using extraction rules and oversight from multiple reviewers. Formal assessment of reproducibility will be a goal of future work.

Large-scale healthcare issues such as COVID-19 or opioid use disorder have dramatic adverse effects on personal and global health. The efforts of thousands of researchers worldwide working on a health problem will be sped up if their collective research output was better organised and extracted into a user-friendly database. This pilot study demonstrated the feasibility to accomplish this task. Treating results like units of information provides freedom to assemble and integrate information from multiple manuscripts in a logical manner. This method of extraction using parent-child relationships and automated linkages facilitates integration of information and makes entry of data easier, especially by less experienced reviewers. We conclude that this system provides a platform to scale up extraction of information on very large topics to be managed by multiple individuals residing in diverse locations.

Contributors All authors contributed equally.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests MB and DK have ownership interest in Plomics. SL, IF, CT-L, SR and JP declare no financial or competing interests.

Patient consent for publication Not required.

Ethics approval This study does not involve human participants.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially,

and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD


David Krag <http://orcid.org/0000-0001-5355-5999>

REFERENCES

- 1 Müller H-M, Van Auken KM, Li Y, *et al.* Textpresso central: a customizable platform for searching, text mining, viewing, and curating biomedical literature. *BMC Bioinformatics* 2018;19:94.
- 2 Simon C, Davidsen K, Hansen C, *et al.* BioReader: a text mining tool for performing classification of biomedical literature. *BMC Bioinformatics* 2019;19:57.
- 3 Subramanyam RV. Art of reading a journal article: Methodically and effectively. *J Oral Maxillofac Pathol* 2013;17:65–70.
- 4 Linzer M, Mercado A, Hupart KH. Role of a medical Journal Club in residency training. *Academic Medicine* 1986;61:471–3.
- 5 Eden J, Levit L, Berg A. *Finding what works in health care: standards for systematic reviews*. Washington (DC), 2011.

© 2022 Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ. <http://creativecommons.org/licenses/by-nc/4.0/>This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>. Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.

Characterising the health and social care segment of the BCS (The Chartered Institute for IT) membership and their continuing professional development needs

Katie MacLure ¹, Andrew MacLure,² Sharon Levy ³, Wendy Dearing⁴

To cite: MacLure K, MacLure A, Levy S, *et al*. Characterising the health and social care segment of the BCS (The Chartered Institute for IT) membership and their continuing professional development needs. *BMJ Health Care Inform* 2022;**29**:e100504. doi:10.1136/bmjhci-2021-100504

Received 26 October 2021
Accepted 20 January 2022

ABSTRACT

Objectives The aim of this study was to identify and characterise the health and social care membership of the British Computer Society (BCS), an international informatics professional organisation, and to determine their ongoing development needs.

Methods A prepiloted online survey included items on professional regulatory body, job role, work sector, qualifications, career stage, BCS membership (type, specialist group/branch activity (committees, event attendance)), use of BCS.org career planning/continuing professional development (CPD) tools, self-reported digital literacy and other professional registrations. The quantitative data were analysed using descriptive statistics in JASP V.0.9.2 to report frequencies and correlations.

Results Responses were received from 152 participants. Most were male (n=103; 68%), aged 50–59 years (n=41; 28%), working in England (n=107; 71%) with master's or honours degrees (n=80; 53%). Most were either new (5 years or less; n=61; 40%) or long-term members (21 years or more; n=43; 28%) of BCS. Most were not interested in health specialist groups (n=57; 38%) preferring non-health specialist groups such as information management (n=54; 37%) and project management (n=52; 34%).

Discussion This is the first paper to characterise the health and social care membership of an IT-focused professional body and to start to determine their CPD needs. There are further challenges ahead in curating the content and delivery.

Conclusion This study is the starting point from which members' CPD needs, and ongoing interest, in being recognised as health and social care professional members, can be acknowledged and explored. Further research is planned with the participants who volunteered to be part of designing future CPD content and delivery.

INTRODUCTION

The British Computer Society (BCS), The Chartered Institute for IT, has a long and distinguished history since it was established in 1957 with a membership over 60 000 across 150 countries.¹ The royal charter made the

Summary

What is already known?

- ▶ British Computer Society (BCS), The Chartered Institute for IT, does not interrogate membership data to determine which of its members identify as health and social care professionals.
- ▶ Therefore, BCS cannot fulfil its responsibility to identify members' relevant continuing professional development (CPD) needs and offers the right opportunities to support their career aspirations.

What does this paper add?

- ▶ This study has identified and characterised the BCS membership segment who self-identified as health and social care professionals and articulated their CPD needs, and ongoing interest, in being recognised as specialists.
- ▶ Further research is planned with the participants who volunteered to be part of ongoing research designing future CPD content and modes of delivery.

BCS a charity 'responsible for raising the standards of IT education, professionalism, ethics and practice' while 'making IT good for society'. Built on five pillars of: (1) sharing expertise, (2) improving education, (3) influencing practice, (4) driving standards and (5) supporting careers, its membership is now drawn from professions as diverse as the technologies which underpin society including health and social care.²

During the COVID-19 pandemic, BCS ran a campaign to celebrate IT professionals as 'val workers' keeping society connected and informed.³ Efforts to manage COVID-19 outbreaks relied on advanced coordinated technologies; the health data scientists and bioinformaticians used digital analytics tools; ordinary citizens relied on digital tools and connectivity for work and education and the health and social care professionals



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Independent Research Consultant, Aberdeen, UK

²Independent Research Associate, Aberdeen, UK

³Usher Institute, University of Edinburgh, Edinburgh, UK

⁴Institute of Management and Health, University of Wales Trinity Saint David, Lampeter, UK

Correspondence to

Dr Katie MacLure;
katiemacLure@outlook.com

transformed their practice while honing their digital literacy to continue and offer optimal (digital) healthcare services.⁴

The Topol Review, published in 2019, focused on 'preparing the healthcare workforce to deliver the digital future'.⁵ Building a digitally ready workforce (BDRW) has been an ongoing strategy for the National Health Services (NHS) across the devolved home nations of the UK (England, Northern Ireland, Scotland and Wales) and increasingly considered for social and care workers too. The review proposed three principles: (1) that patients should be partners in decisions about their health aided by health technologies; (2) that the healthcare workforce needs expertise and guidance to evaluate new technologies and (3) that adoption of new technologies should give health and care professionals 'the gift of time to care'.⁵ Three technologies were specifically mentioned: (1) artificial intelligence (AI); (2) genomics and (3) digital medicine. The review emphasised the importance of a digitally competent health and social care workforce which understands data-driven technologies and is 'digitally confident, digitally aware and digitally literate'. It described new disciplines that were likely to emerge such as higher specialist scientists, knowledge management, AI and robotics engineering, digital health technicians, bioinformaticians and digital technologists.⁵

In anticipation of, and catering for, the learning needs of an emerging workforce, the BCS planned an internal audit to articulate the provision and needs of current members who work in the health and social care arena. The main objective was to identify appropriate learning scaffolding frameworks and provision of 'in house' continuing professional development (CPD) content, which fit the lifelong learning ethos. However, it became clear at an early stage that the organisation does not have, nor is it set to retrospectively collect, data on professional roles or sectors of its membership. It is, therefore, unaware which of their members identify as health and social care professionals. These data are critical in understanding professional learning needs and how to address them.

A 2020 scoping review of 1.5 million registrants identified 32 healthcare professional job titles in the UK.⁶ Each associated with one of the nine regulatory bodies each of which has a different length of CPD cycle (General Optical Council refers to continuing education and training (CET) rather than CPD) ranging from 1 year to 5 years.⁶

An earlier 2019 report, prepared by 'The Interprofessional CPD and Lifelong Learning UK Working Group', identified five principles for CPD and lifelong learning for the health and social care sector.⁷ Principle 1 stated that it would be each person's responsibility and be made possible and supported by their employer; principle 2 stated that it would benefit service users; principle 3 stated that it would improve the quality of service delivery; principle 4 stated that it would be balanced and relevant and finally, principle 5 stated that it would be

recorded and show the effect on each person's area of practice. However, little is included regarding digital (n=0) or informatics (n=0) or technology (n=2) but it calls on professional bodies and trade unions, employers and 'the wider system' to promote CPD to improve the quality of service delivery.⁷

In contrast, a most recent commissioned report published in *The Lancet* considered the future of health and care service post-COVID-19, although 64 pages in length, featured many of these key terms numerous times: digital (n=74), informatics (n=0), technology (n=86) and health (n=1539), social (n=251) and care (n=954).⁸ The report names: Health Education England and the Department of Health and Care; National Health Service Education for Scotland; Health Education and Improvement Wales and Northern Ireland Department of Health responsible for health workforce planning.⁸

There are key skills and competencies frameworks for health and care⁹⁻¹⁵ which have started to include variations on technical efficiency, informatics competence or similar. It may still take a leap of faith to compare, combine or critically appraise such frameworks against the BCS SFIAplus V.7, a task which is outwith the scope of this study.^{16 17} The Skills Framework for the Information Age (SFIA) which, being generic, may lack alignment given health (n=0), social (n=0) and care (n=0) do not feature in SFIAplus.^{16 17}

Given reports that the health and social care professions account for almost 1 in 10 jobs in the UK¹⁸ and in the aftermath of COVID-19 the rapid digitisation of the sector, the BCS, The Chartered Institute for IT, needs to act now. BCS has a responsibility to identify and engage those working with digital health or ehealth or technology enabled care or with health informatics interests and recognise the potential for hybrid career paths which may have specialised CPD needs.¹⁹

Aim of study

Therefore, the aim of this study was to characterise the health and social care membership of BCS and to determine their CPD needs.

METHODS

Design and methods

A quantitative cross-sectional online survey was designed based on a literature review and interviews with key stakeholders (36 representatives of health and social care professions, BCS members, BCS staff).

Setting

The BCS, The Chartered Institute for IT, is the UK's professional body for computing including health and care informatics. The membership represents a broad spectrum of IT professionals but does not currently collect data on employment sector so cannot target relevant communications.

Inclusion and exclusion criteria

The survey was open to all BCS members who self-identified as health or social care professionals.

Data collection tools

The survey was reviewed for face and content validity within the research team before piloting with five key stakeholders who had previously taken part in a related interview. The survey was hosted online by BCS and shared with the whole membership by email inviting participation by anyone self-identifying as a health or social care professional. Two reminders were sent. The link to the survey was also promoted in newsletters, on social media and with BCS specialist groups.

Questions asked were related to: professional regulatory body, job role or title, work sector, highest qualification, career stage, BCS membership (type, years since enrolled, specialist group interests and branch activity (committees, event attendance)) and use of BCS.org career planning and CPD tools, self-reported digital literacy and other professional registrations. An open text question, which is reported elsewhere, asked what CPD content the sector wanted BCS to provide. The survey was anonymous but participants had the opportunity to opt in to further involvement including: to be recognised by BCS as a health and social care professional, take part in a follow-up interview and join a consensus panel to design/decide on BCS CPD provision for the health and social care membership.

Data collection

The survey was open from 13 January to 16 March 2021. Completion of the survey was taken as informed consent.

Data analysis

Only the quantitative data from the survey are reported in this article. These were analysed using descriptive statistics in JASP V.0.9.2, the open source statistical programme, to report frequencies and correlations.

RESULTS

Responses were received from 152 participants which is a tiny proportion of the 60 000 international membership. As per [table 1](#), most were male (n=103; 68%) with the highest proportion in the 50–59 years age bracket (n=41; 28%) and working in England (n=107; 71%). This educated workforce reported their highest qualification gained as foundation degree level (n=37; 24%), master's or honours degree level (n=80; 53%) or doctoral level (n=19; 13%). Many were also members or registered with one or more professionally recognised organisations including BCS Federation of Informatics Professionals (FED-IP; n=23; 16%) or the Institute of Engineering/Chartered Engineer (n=18; 12%) or Registered IT Technician (n=16; 11%). However, more than half (n=81; 55%) were not. The majority considered themselves to be mid-career (n=64; 42%) with few early in their career

Table 1 Demographics and BCS membership (N=152)

Do you identify as?	n (%)
Male	103 (68)
Female	45 (30)
Prefer not to say	3 (2)
Which age group are you in?	
Under 20 years	0 (0)
20–29 years	8 (5)
30–39 years	22 (15)
40–49 years	28 (19)
50–59 years	41 (28)
60–69 years	29 (20)
70 years or over	20 (14)
Which country do you mainly work in?	
England	107 (71)
Wales	23 (15)
Scotland	9 (6)
Northern Ireland	3 (2)
Other: UK (n=3), Hong Kong (n=2), Luxembourg, Sri Lanka, Singapore, international bodies	9 (6)
Which level is your highest qualification?	
Doctorate	19 (13)
Master's or honours degree/postgraduate certificate/diploma/NVQ5/SVQ5	80 (53)
HNC/D or foundation/ordinary/bachelor's degree/NVQ4/SVQ4	37 (24)
Scottish highers/advanced highers/A levels/National 5/NVQ3/SVQ3	7 (5)
GCSE/standard grade/National 4/NVQ2/SVQ2 or equivalent	6 (4)
Other: BA (Hons) plus FCCA, M.B.B.S., CISSP	3 (2)
Are you a member or registered with any of the following?	
FED-IP	23 (16)
IEng/CEng	18 (12)
RITTech	16 (11)
FCI	12 (8)
CHIME	11 (8)
HIMSS	7 (5)
Other: InstRE, FCyBS, European Resuscitation Council, IAHSI, Chartered Management Institute, IEEE, BCS Elite IT Leaders Forum, IHM, IMIA, Institute of Leadership and Management, Institution of Civil Engineers, IAP	15 (10)
None of the above	81 (55)
In terms of your career, do you consider yourself to be?	
Early career/newly qualified/new entrant	20 (14)
Mid-career	64 (42)
Looking towards retirement	36 (24)

Continued

Table 1 Continued

Retired	32 (21)
Which level of BCS membership do you have?	
Professional (MBCS)	67 (44)
Chartered IT professional (MBCS CITP)	23 (15)
Associate (AMBCS)	22 (15)
Chartered fellow (FBCS CITP)	13 (9)
Fellow (FBCS)	10 (7)
Student	9 (6)
Affiliate	8 (5)
Is that through?	
Individual membership	113 (75)
Organisational membership	37 (25)
How long have you been a BCS member?	
5 years or less	61 (40)
6–10 years	17 (11)
11–15 years	17 (11)
16–20 years	14 (9)
21 years or more	43 (28)

BCS, British Computer Society; FED-IP, Federation of Informatics Professionals; IEng/CEng, Institute of Engineering/Chartered Engineer; RITTech, Registered IT Technician.

(n=20; 14%). The survey attracted participation from a sizeable group of retired IT professionals (n=32; 21%) and those looking towards retirement (n=36; 24%). Most were professional members of the BCS (MBCS; n=67; 44%) or chartered IT professionals (n=23; 15%); very few were student members of BCS (n=9; 6%). A quarter of the respondents' BCS membership was through their employment organisation (n=37; 25%) with the majority holding individual membership (n=113; 75%). The number of years of membership was dominated by new (5 years or less; n=61; 40%) or long-term membership (21 years or more; n=43; 28%).

In [table 2](#), there was representation from the Nursing and Midwifery Council (n=13; 9%), Health and Care Professions Council (n=8; 5%), the General Medical Council (n=6; 4%) with few responses from the General Dental Council, General Pharmaceutical Council or Social Work England (each n=2; 1%) and Social Care Wales (n=1; 1%) is shown. There was no participation from the General Chiropractic Council, General Optical Council, General Osteopathic Council, Northern Ireland Social Care Council, Pharmaceutical Society of Northern Ireland, Scottish Social Services Council or Scottish Care. A large proportion was not associated with any health and social care regulatory body (n=91; 61%).

Respondents worked in multiple sectors which, for most, were NHS based (n=110; 73%) or corporate IT (n=33; 22%). Although low in numbers, the breadth of sectors was demonstrated with residential and day care for older people (n=5; 3%), adults (n=3; 2%) and children

Table 2 Regulatory bodies and employment sectors (N=152)

Regulatory body	n (%)
Nursing and Midwifery Council	13 (9)
Health and Care Professions Council	8 (5)
General Medical Council	6 (4)
General Dental Council	2 (1)
General Pharmaceutical Council	2 (1)
Social Work England	2 (1)
Social Care Wales	1 (1)
General Chiropractic Council	0 (0)
General Optical Council	0 (0)
General Osteopathic Council	0 (0)
Northern Ireland Social Care Council	0 (0)
Pharmaceutical Society of Northern Ireland	0 (0)
Scottish Social Services Council	0 (0)
Scottish Care	0 (0)
None of the above	91 (60)
Other: FEDIP/UKCHIP (n=4), UK Council for Psychotherapy (n=2), IAHSI (n=2), BACP (n=2), ISC (n=2), NWIS (n=2), Society and College of Radiographers, NCS, Public Health, CPCAB, Care Quality Commission, Association of Clinical Biochemists, Institute of Biomedical Science, IHM, European Resuscitation Council, EFMI, IMIA, BCS, NHS Trust, ACCA, ISACA, IAPP, SABSA Institute	28 (18)
Which sectors do you or did you work in?	
NHS	110 (72)
Corporate IT	33 (22)
Academia/education	24 (16)
Research/consultancy	23 (15)
Primary care	23 (15)
Secondary care	22 (15)
Local government	20 (13)
Voluntary sector	20 (13)
Freelance/independent	18 (12)
Industry	15 (10)
Third sector	14 (9)
National government	12 (8)
Intermediate care	10 (7)
Emergency care	9 (6)
Social work	9 (6)
Performance	8 (5)
Other community-based support services	8 (5)
Residential care (adults)	6 (4)
Care at home	6 (4)
Residential care (older people)	5 (3)
Residential care (children)	3 (2)

Continued

Table 2 Continued

Day care services (adults)	3 (2)
Housing support	2 (1)
Day care services (children)	2 (1)
Day care services (older people)	1 (1)
Other: ExE for CQC—adult social care, civil service—defence primary healthcare, social care system software supplier, project management and business analysis, mental healthcare, social care membership body, consultancy, government departments, NIHR and HDRUK	13 (9)

NHS, National Health Services.

(n=3; 2%) as well as housing support (n=2; 1%) and care at home (n=6; 4%).

Survey respondents were associated with a range of BCS specialist groups and branch committees (table 3). While primary care was the most frequently indicated (n=44; 29%), a larger proportion was not interested in any of these specialist groups (n=57; 38%). A similar proportion was interested in non-health specialist groups such as information management (n=54; 37%) and project management (n=52; 34%). Overall, although participants self-identified as health and social care professionals, many indicated more interest in non-health specialist groups.

In relation to branch committee membership, more than a third were unaware of the opportunity (n=52; 34%) with just over a fifth either a current (n=21; 14%) or past (n=11; 7%) branch committee member.

Table 4 gauges the digital literacy of the participants which in most topic areas is 'confident and capable' with the exception of 'creation, innovation and research' which dips to 'can use' (n=52; 36%) and awareness 'know' (n=26; 18%). There is still a sizeable proportion who describe themselves as an 'expert user' particularly noticeable for the topic area 'information, data and content' (n=35; 23%) and 'technical proficiency' (n=29; 19%).

When asked which recent BCS Health and Care webinar titles most appealed (table 5), participants found 'data enabled technologies and services in health and social care' most appealing (n=57; 38%). This was the case for both retired and looking towards retirement (n=24/68; 35.3%) and other earlier career stages (n=33/84; 39.3%). Second most popular was 'building a digitally ready workforce in health and social care' (n=46; 34%). While the appeal of 'ethics and AI' and 'co-creating digital medicine technologies' were unclear, participants found 'a framework for genomic leadership' least appealing (n=74; 63%). Again, this 'least appealing' topic was the case for retired and looking towards retirement (31/68; 45.6%) and earlier career stages (43/84; 51.2%).

Table 3 Interest in health and other BCS specialist groups and branch committees (N=152)

Which of these existing health and other BCS specialist groups are you interested in or signed up to follow?	n (%)
Primary Care	44 (29)
National Mental Health	25 (16)
Health Nursing	22 (14)
Health & Care Wales	17 (11)
Health & Care Northern	15 (10)
Health & Care Scotland	10 (7)
None of the above	57 (38)
Other: BCS Women (n=2), AI (n=2), Health Informatics, Primary Care, Health Nursing, Health London & South East, Health London, GP Specialist Group, Health Informatics, Social Care, Allied Health Professions, Acute, Genomics, Clinical Best Practise, Telemedicine, SGAI, District Nursing and community care, London Medical	12 (8)
Are there any other existing BCS specialist group areas you are interested in or signed up to follow?	
Information Management	54 (36)
Project Management	52 (34)
Learning & Development	39 (26)
Ethics, Law & Diversity in IT	35 (23)
Business & Consultancy	33 (22)
Strategy & Architecture	32 (21)
Future of Computing	30 (20)
Security	29 (19)
Software Development	28 (18)
History of Computing	17 (11)
None of the above	14 (9)
Other: BCS Women, AI, Software Testing, Data Scientist, IRMA, Elite IT, Digital Informatics and Data Analytics (BI, AI and Machine Learning), North London, Data Management, Central London, SGAI, Artificial Intelligence	14 (9)
Are you or have you ever been a member of your local branch committee?	
No—not interested	67 (44)
No—wasn't aware of opportunity	52 (34)
Yes—currently	21 (14)
Yes—in the past	11 (7)

BCS, British Computer Society.

DISCUSSION

This is the first paper to characterise the health and social care membership of BCS and to start to determine the CPD needs of this diverse population. From the results, participants form a 'digitally confident, digitally aware and digitally literate'⁵ group meeting the target competencies identified in the Topol Review,⁵ Karas *et al's* review,⁶ Broughton *et al's* report⁷ and the competencies frameworks from across the health and social care professions and the home nations.^{9–15} It is clear that the trajectory is towards BDRW which may have gained momentum

Table 4 What is your level of digital literacy in relation to the topic areas listed below?

Digital literacy topic area	I know there are many related digital tools and technologies	I can use related digital tools and technologies	I am confident and capable in the use of a wide range of related digital tools and technologies	I am an expert user and take a lead in modelling and promoting the use of a wide range of related specialist digital tools and technologies
Information, data and content (n=151)	11 (7)	29 (19)	76 (50)	35 (23)
Teaching, learning and self-development (n=149)	13 (9)	42 (28)	74 (50)	20 (13)
Communication, collaboration and participation (n=150)	11 (7)	40 (27)	78 (52)	21 (14)
Creation, innovation and research (n=144)	26 (18)	52 (36)	47 (33)	19 (13)
Technical proficiency (n=149)	19 (13)	42 (28)	59 (40)	29 (19)
Digital identity, well-being, safety and security (n=149)	20 (13)	42 (28)	65 (44)	22 (15)

Most frequent highlighted in bold

during the COVID-19 pandemic.^{8 20–23} Whether that momentum of improving digital competency can be continued post COVID-19, with a workforce which has been overwhelmed during the pandemic, remains to be seen. It should be also be noted that the three technologies highlighted in the Topol Review as important for the future of health and social care, namely AI, genomics and digital medicine, were the least popular webinar topics for this group of respondents.⁵

BCS do not collect data on professional roles or sectors. They do not know which of their members identify as health and social care professionals, so consideration needs to be given to inviting the membership to share details which can be the foundation for targeting relevant CPD opportunities. Not only would that provide insight into the 37 listed professions^{6,7} but also into the relevant regulatory and professional bodies so BCS can complement rather than replicate their CPD offering.

Table 5 Which of these example webinar event titles most appeals to you?

Webinar titles	Mean	1—most appealing	2	3	4	5—least appealing
Data Enabled Technologies and Services in Health and Social Care (n=138)	1.96	57 (38)	41 (27)	31 (20)	7 (5)	2 (1)
Building a Digitally Ready Workforce in Health and Social Care (n=134)	2.25	46 (34)	37 (28)	29 (22)	16 (12)	5 (4)
Ethics of Artificial Intelligence and Autonomous systems in Health and Social Care (n=135)	2.88	29 (21)	28 (21)	28 (21)	30 (22)	20 (15)
Co-creating Digital Medicine Technologies with Health and Social Care Staff (n=127)	2.96	20 (16)	25 (20)	34 (27)	36 (28)	12 (9)
A Framework for Genomic Leadership across Care Sectors (n=118)	4.37	4 (3)	6 (5)	6 (5)	28 (24)	74 (63)

Where clear, most frequent highlighted in bold

The recent *Lancet* paper⁸ names: Health Education England and Department of Health and Care; National Health Service Education for Scotland; Health Education and Improvement Wales and Northern Ireland Department of Health as responsible for health workforce planning. This highlights further opportunities for meaningful collaboration to grow the range of CPD on offer. Globally, the challenge has been highlighted by the Organisation for Economic Co-operation and Development (OECD) in their 2021 report into 'Empowering the Health Workforce'.¹⁶ The OECD states that, 'To meet the current demand for digital upskilling, the CPD and other professional training schemes should become a shared responsibility between employers, professional organisations, and ministries of health'.¹⁶

It may still take a leap of faith to compare, combine or critically appraise the many frameworks^{9–15} against the BCS SFIaplus V.7^{17 18} but this task is outwith the scope of this study. There are many other players in the CPD arena, such as the NHS Digital Academy^{24 25} and, for this mainly highly educated group of professionals, wider options provided by over 80 MSc courses in health data sciences, analytics and informatics.²⁰ Certainly, OECD notes that 'the pace of changes has been particularly slow with regard to whether and how the CPD and other on the job training include digital health content'.¹⁶

But, the obstacle is that BCS currently do not know how to meaningfully identify and support their health and social care professional membership with their CPD, CET or lifelong learning needs. It was interesting to note and useful for people organising events and content that participants from all career stages showed commonality in the webinar topics which most and least appealed to them. It is also unclear from the results whether the health and social care professional really understands who and what the BCS is, the purpose of BCS, how BCS can support the breadth of health and social care professionals and what it can offer. If BCS is to support the hybrid careers of health and social care professionals by providing relevant CPD, it must first identify the segment of the membership.

With the BCS FED-IP reporting six themes in their 'Becoming the Profession'²⁶ as: (1) Recognition; (2) CPD; (3) Accreditation, Education and Training, (4) Career Guidance and Support, (5) Networking and (6) Simplifying the Landscape, there is clear alignment with the results of this report plus interest and willingness to explore this complexity.²⁷ However, there is a lot more to be done in engaging meaningfully with the health and social care professionals and their communities of practice, to optimise across the relevant organisations the CPD offering to each is best situated to provide.

Limitations

The participants self-identified as health and social care professionals but many were not registered with a regulatory body. Moreover, the characteristics of the sample are very different to the population of mainly female staff

working in health and social care settings. This raises questions around shared understanding of whom among the membership fit the BCS target group. This lack of a denominator also makes it impossible to calculate a response rate but clearly higher participation would be helpful in achieving generalisability. If BCS were to give the applicant the opportunity to share their professional and role details on registration or during an annual review, the role BCS could fulfil with regard to CPD would be much simpler to follow-up and action. A strength of the study is the adoption of the Consensus-Based Checklist for Reporting of Survey Studies.²⁸

CONCLUSION

In conclusion, BCS has a responsibility to provide its members with the CPD content that is relevant to their career path and aspirations. To date, BCS has not been able to target the health and social care segment of the membership. This study has identified and characterised that segment of professionals who self-identified, and have indicated, their CPD needs and ongoing interest in being recognised by BCS as health and social care professionals with BCS membership. Further research is planned with the participants who volunteered to be part of ongoing research for designing future CPD content and delivery.

Twitter Katie MacLure @katiemaclure

Acknowledgements The authors gratefully acknowledge those who initiated the design of the research, helped to pilot the study and took part in the face and content validity and staff at British Computer Society headquarters who hosted and distributed the survey. Most of all the authors thank those who took the time to participate in the study.

Contributors SL and WD: conceived and gained funding, kept oversight of the project, commented on design of data collection and analysis and critically revised versions of the paper. KM and AM: designed and conducted data collection and analysis and wrote the initial draft of the paper, which was revised by KM. All authors approved the final version for which KM acts as guarantor for the conduct and publication of its' content.

Funding This research was funded by the BCS Health and Care Executive.

Competing interests KM, SL and WD are members of the British Computer Society Health and Care Executive. Otherwise, they have no conflicts of interest to declare.

Patient consent for publication Not applicable.

Ethics approval This study involved human participants but ethical approval was not explicitly sought as it is included in the legal and privacy notices for British Computer Society (BCS) members (<https://www.bcs.org/legal-and-privacy-notices/>). The BCS Data Privacy Notice on 'how we use your personal data' includes provision of 'surveys, information about authors' awards and events, offers and promotions, related to the products and/or services'. The survey was reviewed by BCS community team and BCS Health and Care Executive. Voluntary completion of the survey was taken as participant informed consent. Participants gave informed consent to participate in the study before taking part. In addition, the study was deemed retrospectively, to be a service evaluation exempt from Research Ethics Committee review.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is

properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Katie MacLure <http://orcid.org/0000-0003-0686-948X>

Sharon Levy <http://orcid.org/0000-0001-6208-6928>

REFERENCES

- BCS. BCS website – about us, 2021. Available: <https://www.bcs.org/more/about-us/> [Accessed 16 Aug 2021].
- BCS. BCS homepage - The professional body for computing, 2021. Available: <https://www.bcs.org/> [Accessed 16 Aug 2021].
- Hamilton J. Recognising vITal workers. *ITNOW* 2020;62.
- Fagherazzi G, Goetzing C, Rashid MA, et al. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. *J Med Internet Res* 2020;22:e19284.
- Topol E. The Topol review. preparing the healthcare workforce to deliver the digital future, 1-48, 2019. Available: <https://topol.hee.nhs.uk/wp-content/uploads/2019/02/Topol-Review-FAQs.pdf> [Accessed 15 Aug 2021].
- Karas M, Sheen NJL, North RV, et al. Continuing professional development requirements for UK health professionals: a scoping review. *BMJ Open* 2020;10:e032781.
- Broughton W, Harris G, on behalf of the Interprofessional CPD and Lifelong Learning UK Working Group. *Principles for continuing professional development and lifelong learning in health and social care. Bridgwater: College of Paramedics*, 2019. <https://www.collegeofparamedics.co.uk/publications/principles-for-cpd>
- Anderson M, Pitchforth E, Asaria M, et al. LSE–Lancet Commission on the future of the NHS: re-laying the foundations for an equitable and efficient health and care service after COVID-19. *The Lancet* 2021;397:1915–78.
- NHS England. A health and care digital capabilities framework, 2018. Available: <https://www.hee.nhs.uk/sites/default/files/documents/Digital%20Literacy%20Capability%20Framework%202018.pdf> [Accessed 15 Aug 2021].
- Agenda for Change Project Team. The NHS knowledge and skills framework (NHS KSF) and the development review process. Department of health, 2004. Available: https://webarchive.nationalarchives.gov.uk/ukgwa/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4090861.pdf [Accessed 17 Aug 2021].
- NHS Scotland (2021) key skills framework. Available: <http://www.ksf.scot.nhs.uk/> [Accessed 17 August 2021].
- NHS Wales (2004) key skills framework. Available: <http://www.wales.nhs.uk/sites3/Documents/433/KSF04088218.pdf> [Accessed 17 Aug 2021].
- Brice S, Almond H. Health professional digital capabilities frameworks: a scoping review. *J Multidiscip Healthc* 2020;13:1375–90.
- NHS England. A digital framework for allied health professionals, 2019. Available: <https://www.england.nhs.uk/wp-content/uploads/2019/04/a-digital-framework-for-allied-health-professionals.pdf> [Accessed 18 Aug 2021].
- Li S, Bamidis PD, Konstantinidis ST, et al. Setting priorities for EU healthcare workforce IT skills competence improvement. *Health Informatics J* 2019;25:174–85.
- Socha-Dietrich K. Empowering the health workforce to make the most of the digital revolution, 2021. Available: <https://www.oecd.org/health/health-systems/Empowering-Health-Workforce-Digital-Revolution.pdf> [Accessed 15 Nov 2021].
- BCS. Browse SFIaplus, 2020. Available: <https://mybcs.bcs.org/career-development/browse-sfiaplus/> [Accessed 16 Aug 2021].
- BCS. Skills framework for an information age (SFIaplus v7), 2020. Available: <https://sfiaplus.bcs.org/SfiaPlus> [Accessed 16 Aug 2021].
- The King's Fund. Overview of the health and social care workforce. Available: https://www.kingsfund.org.uk/projects/time-think-differently/trends-workforce-overview?utm_source=twitter&utm_medium=social&utm_term=socialshare [Accessed 29 Sep 2021].
- Prospects (2021) UK masters courses in informatics. Available: <https://www.prospects.ac.uk/postgraduate-courses/masters-degrees?keyword=informatics&featuredCourses=138851%2C125812&page=0> [Accessed 18 Aug 2021].
- He W, Zhang ZJ, Li W. Information technology solutions, challenges, and suggestions for tackling the COVID-19 pandemic. *Int J Inf Manage* 2021;57:102287.
- Vargo D, Zhu L, Benwell B, et al. Digital technology use during COVID -19 pandemic: A rapid review. *Hum Behav Emerg Technol* 2021;3:13–24.
- Bokolo AJ. Application of telemedicine and eHealth technology for clinical services in response to COVID-19 pandemic. *Health Technol* 2021;11:359–66.
- Higher Education England (2021) Digital Health Academy. Available: <https://www.hee.nhs.uk/our-work/nhs-digital-academy> [Accessed 18 Aug 2021].
- The Knowledge Network (2021) NMAHP digital health and care leadership programme. Available: <http://www.knowledge.scot.nhs.uk/enmahpleadershipcommunity/resources.aspx> [Accessed 18 Aug 2021].
- Federation of Informatics Professionals. Becoming the profession, 2021. Available: <https://www.fedip.org/becoming-the-profession> [Accessed 17 Aug 2021].
- Mbunge E, Akinnuwesi B, Fashoto SG, et al. A critical review of emerging technologies for tackling COVID -19 pandemic. *Hum Behav Emerg Technol* 2021;3:25–39.
- Sharma A, Minh Duc NT, Luu Lam Thang T, et al. A consensus-based checklist for reporting of survey studies (cross). *J Gen Intern Med* 2021;36:3179–87.

© 2022 Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ. <http://creativecommons.org/licenses/by-nc/4.0/>This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>. Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.