

Background Clinical prediction models are used for different purposes, but purpose-specific validation is not usually carried out. The ability of a model to discriminate between true positives and false positives has applications in clinical decision making, screening, and service evaluation. The calibration (goodness-of-fit) of a model is a key indicator of how well a model's predicted outcomes reflect those actually observed. Initial validation of models usually includes assessment of these features but re-evaluation over time might not be performed.

EuroSCORE is an adult cardiac surgery risk model which has been in use since 1998. It predicts in-hospital mortality and is used for clinical decision making and service evaluation. It is widely acknowledged to have demonstrated 'calibration drift', but this has not been formally evaluated in the UK population.

Methods We assessed the performance of EuroSCORE in the Central Cardiac Audit Database (CCAD), covering all NHS cardiac procedures in the UK. Discrimination was tested using the area under the Receiver Operator Characteristic (ROC) curve (AUC). Calibration was assessed with the Hosmer-Lemeshow goodness of fit test. In addition, we developed new models with longer-term outcomes using the data, and tested year-on-year model performance.

Results A total of 399,314 eligible procedures from 1st April 1998 to 31st March 2011 were included in the analysis. Assessing the discrimination of EuroSCORE by financial year showed consistency across the period (AUC values ranging from 0.788 to 0.818). Model calibration, however, drifted considerably with a cumulative mortality over-estimate of 10,801 deaths by the end of the period (increasing from 147 over-estimated deaths in 1998 to 1,500 in 2010). This represented a predicted overall mortality rate of 6.0% compared with the observed rate of 3.4%. We will also present findings relating to year-on-year performance of a panel of models tailored to longer-term outcomes in specific procedures.

Conclusion Models that retain accurate discrimination while undergoing calibration drift may be implemented in settings for longer than is appropriate. A model that maintains good discrimination may be useful in a subset of scenarios, but for most purposes good calibration is also crucial. For models developed for multiple applications, purpose-specific validation and recalibration should be considered. Model performance should be appraised in context and not by indicators in isolation.

OP72 PROSPECTIVE EXTERNAL VALIDATION OF RISK PREDICTION MODELS FOR ACUTE TRAUMATIC BRAIN INJURY IN UK CRITICAL CARE UNITS: THE RAIN STUDY

doi:10.1136/jech-2012-201753.072

¹DA Harrison, ¹KA Griggs, ²M Gomes, ³DK Menon, ¹KM Rowan. ¹*Clinical Trials Unit, Intensive Care National Audit & Research Centre (ICNARC), London, UK;* ²*Department of Public Health & Policy, LSHTM, London, UK;* ³*School of Clinical Medicine, University of Cambridge, Cambridge, UK*

Background Acute traumatic brain injury (TBI) is the leading cause of death and disability in adults aged under 40 years. Statistical models have been developed to predict the risk of mortality or unfavourable outcome (death or severe disability) at six months following acute TBI but to date these risk prediction models have only been validated using existing data sources. The Risk Adjustment In Neurocritical care (RAIN) Study aimed to validate these risk prediction models among adults with acute TBI admitted to UK critical care units.

Methods Ten risk prediction models were identified: four for mortality at six months (the Hukkelhoven model and IMPACT Core, Extended and Lab models); and six for unfavourable outcome at six months (as mortality plus CRASH Basic and CT models). Risk factor data were collected from 67 UK critical care units (including 90% of regional neuroscience centres) from August 2009 to March 2011. Patients were followed up to six months for mortality by

linkage with death registration and unfavourable outcome using the Glasgow Outcome Scale (Extended) administered by postal or telephone questionnaire.

The risk prediction models were validated for calibration (c index), discrimination (Hosmer-Lemeshow test and Cox calibration regression) and overall fit (Brier score). Missing data were handled with multiple imputation.

Results Data were collected for 2,975 eligible patients admitted to critical care following acute TBI. 97% of patients were followed-up for mortality and 81% for unfavourable outcome at six months. Following multiple imputation, mortality and unfavourable outcome at six months were 26% and 57%, respectively. Risk prediction models for mortality at six months had good discrimination (c index 0.75–0.78) and the Hukkelhoven and IMPACT Lab models were well calibrated, although the IMPACT Core and Extended models over-predicted mortality. The models for unfavourable outcome at six months had worse discrimination (c index 0.69–0.71) and all models substantially under-predicted risk of unfavourable outcome. The best performance overall was found for the IMPACT Lab model, which was the most complex model, incorporating laboratory measurements. Models of the next level of complexity (Hukkelhoven, CRASH CT, IMPACT Extended) all performed similarly.

Conclusion Risk prediction models for acute TBI had acceptable discrimination among a large, representative sample of patients admitted to UK critical care units. Calibration was good for mortality but poor for unfavourable outcome, and these models therefore require recalibration for use in this setting.

Friday 14 September 2012

Parallel Session D

Research Methods: Surveys and Use of Routine Data

OP73 EXPLORING IMPACTS OF SURVEY NON-RESPONSE USING RECORD-LINKAGE OF SCOTTISH HEALTH SURVEY DATA (2003 TO 2008)

doi:10.1136/jech-2012-201753.073

¹L Gray, ²G McCartney, ³IR White, ⁴L Given, ¹SV Katikireddi, ¹AH Leyland. ¹*Social and Public Health Sciences Unit, MRC/CSO, Glasgow, UK;* ²*Health Scotland, NHS, Glasgow, UK;* ³*Biostatistics Unit, MRC, Cambridge, UK;* ⁴*Scottish Centre for Social Research, Nat-Cen, Edinburgh, UK*

Background Inference on population health is commonly derived from health survey data, based on the assumption that they are representative of the target communities. Departure from representativeness may weaken external validity leading to biased estimates, with important implications for public health evidence, particularly in relation to prevalence and quantity estimates such as population alcohol consumption. A key aspect determining the extent to which surveys are representative is the level of participant response. Inverse probability weights based on a limited range of demographic variables are usually applied in an attempt to correct for non-response. We aimed to investigate whether weighted estimates of all-cause mortality and mortality from alcohol-related conditions derived from the 2003 Scottish Health Survey (SHeS) – with household response of 67% – reflect those in the population of Scotland.

Methods Baseline observations from SHeS were individually record-linked to mortality data for the 91% of respondents consenting to linkage, and directly age-standardised survey-weighted mortality rates were calculated for the 3117 men and 3980 women aged 20 years and older at interview. Equivalent mortality rates were calculated for the whole of Scotland in the same age group using

population estimates and mortality data contemporaneous with the linked survey data.

Results There were 201 (6.4%) male deaths and 215 (5.4%) female deaths in the 2003 SHeS by the end of 2008. Among men, all-cause mortality was markedly lower in the SHeS sample (918 per 100,000 person-years [95% CI:850–987]) than in the Scottish population (1361 [95% CI:1357–1365]). Figures for women were also highly significantly different (739 [95% CI:682–795] for the SHeS and 928 [95% CI:925–931] for the Scottish population). Alcohol-related mortality was lower in the SHeS sample (38 [95% CI:18–57] in men and 11 [95% CI:0–22] in women) relative to the Scottish population (57 [95% CI:56–58] in men [non-significant] and 25 [95% CI:24–25] in women [significant]).

Conclusion Respondents to the 2003 SHeS differ from the population they are intended to represent, with much lower than expected all-cause mortality in both sexes; alcohol-related mortality rates were somewhat lower than expected suggesting lower alcohol consumption among survey respondents which, if genuine, would lead to inherent underestimation of population consumption levels. Importantly, differences existed despite the application of conventional weighting and age-standardisation methods. Consideration should be given to the levels of resource allocated for increasing survey response and the further development of survey methodology to address the resultant systematic bias in health survey data arising from non-response.

OP74 MEASURING SEXUAL BEHAVIOUR COMES OF AGE: A COMPARISON OF OUTCOMES IN THE 2010 HEALTH SURVEY FOR ENGLAND WITH THE NATIONAL SURVEY OF SEXUAL ATTITUDES AND LIFESTYLES

doi:10.1136/jech-2012-201753.074

¹CH Mercer, ²C Robinson, ²R Craig, ³A Nardone, ¹AM Johnson. ¹*Infection & Population Health, UCL, London, UK*; ²*Health & Wellbeing Group, National Centre for Social Research, London, UK*; ³*HIV/STI Department, Health Protection Agency, London, UK*

Background The Health Survey for England (HSE) is a general health survey administered annually to a probability sample of people living in England. In 2010, the HSE included, for the first time, questions about sexual health, which previously were considered too sensitive for a general health survey. This paper compares the reporting of sexual behaviours by people aged 16–44 in HSE–2010 with data collected by the second British National Survey of Sexual Attitudes and Lifestyles (Natsal–2), Britain's most recent, dedicated national probability survey of sexual behaviour.

Methods In HSE–2010, 8,420 people aged 16–69 were interviewed, of whom 2,911 were aged 16–44. Natsal–2 interviewed 12,110 people aged 16–44 in 1999/2001. HSE–2010 used pen-and-paper self-completion questionnaires for the sexual health questions, while Natsal–2 used computer-assisted personal-interviews including computer-assisted self-interview for the more sensitive questions, including those on sexual health. HSE–2010 used the same question wording developed and piloted for Natsal.

Results Collecting sexual behaviour data was acceptable to HSE–2010 participants with low item non-response (5–10%), albeit slightly higher than in Natsal–2 (<5%). Reported age at first heterosexual intercourse was comparable in the two surveys: median ages of 17 (men) and 16 (women) among those aged 16–24. However, for some very sensitive questions there were lower levels of reporting in HSE–2010 than in Natsal–2: while the proportion reporting same-sex in the last 5 years was similar (2–3%), reporting of ever having same-sex was lower in HSE–2010 for men (2% vs. 5% in Natsal–2). Similarly, the mean number of opposite-sex partners reported in HSE–2010 was a little lower than in Natsal–2, particularly among men: 9.5 vs. 12.7, respectively, in contrast to 5.4 vs. 6.5, respectively, among women. Men in HSE–2010 were also slightly less likely to report being diagnosed with sexually transmitted infection(s) than

in Natsal–2: 8.7% vs. 11.1%, respectively; but this was reported by similar proportions of women: 12% in both surveys.

Conclusion The inclusion of questions on sexual behaviour in HSE–2010 has demonstrated the feasibility and utility of measuring sexual behaviour in general health surveys, albeit in less detail than in a survey focused on sexual behaviour such as Natsal. General health surveys such as the HSE provide a useful vehicle for monitoring sexual risk behaviour more frequently than is possible with decennial Natsal surveys. Health surveys should not shy away from also measuring sexual behaviour, especially in light of the associations between health status and sexual health outcomes of satisfaction, function and behaviour.

OP75 THE RELATIVE IMPORTANCE OF OBJECTIVE AND SUBJECTIVE SOCIO-ECONOMIC AND SCHOOL-BASED SOCIAL STATUS FOR ADOLESCENT HEALTH AND SUBSTANCE USE

doi:10.1136/jech-2012-201753.075

H Sweeting, K Hunt, A Bhaskar. *MRC Social and Public Health Sciences Unit, University of Glasgow, Glasgow, UK*

Background Contrasting with findings from adults and children, most studies of adolescents find little variation in health according to conventional (objective) socio-economic status (SES) measures. Adolescent smoking is patterned by SES, but relationships between SES and drinking are weaker or non-existent. *Subjective* status captures perceptions of relative rank and may also be important for health. Most studies of health and subjective status focus on adults, use subjective SES and explore self-reported health rather than behaviours. However, subjective *school-based* status may be more important than subjective SES for health in adolescence. This study examines the relative importance of objective SES, subjective SES and school-based social status for adolescent self-reported health and health behaviours.

Methods Data were obtained via schools-based self-completion questionnaires in 2010 with follow-up in 2011 when 2,503 (85% of baseline) 13–15 year olds participated. Variables allowing derivation of family affluence were included in 2010; all other data were obtained in 2011. Scottish Index of Multiple Deprivation (SIMD) was derived via postcodes. Pupils rated subjective SES via the MacArthur Scale of Youth Subjective Social Status, a 10-rung ladder with the top representing 'the best off people in Scotland'. Seven ladders asked them to rate various aspects of their own status, compared to their school year-group. Questionnaires also asked about self-rated health, psychological distress (GHQ-12), smoking and drinking. Analyses suggested three subjective school-based social status dimensions: 'peer', 'scholastic' and 'sports'. Objective SES and all social status measures were each collapsed into three categories for inclusion in logistic regression analyses which were conducted on those with full data (N=1,819) on these measures.

Results Correlations between objective SES and all subjective status measures were weak. In preliminary multivariate logistic regression analyses, adjusted for gender and age, family affluence was not associated with health, smoking or drinking and deprivation was not associated with health. However, each subjective school-based status measure was associated with both health and behaviours. For example, odds (95% confidence intervals) of fair/poor self-rated health among those 'low' compared with 'high': family affluence 1.1 (0.68–1.81); SIMD 1.23 (0.86–1.76); subjective SES 1.42 (0.97–2.08); subjective 'peer' status 1.73 (1.20–2.50); 'scholastic' 2.93 (2.01–4.27); 'sports' 2.93 (1.98–4.35). Odds of ever smoking among those low (vs. high): family affluence 1.43 (0.93–2.21); SIMD 2.28 (1.67–3.13); subjective SES 1.20 (0.84–1.71); 'peer' 0.30 (0.21–0.42); 'scholastic' 11.80 (8.05–17.29); 'sports' 2.00 (1.41–2.84).

Conclusion Subjective school-based social status is more important for adolescent health and substance use than either objective or subjective SES measures.