

Article

Comparison of Familial, Polygenic and Biochemical Predictors of Mortality

John B. Whitfield¹, Lucía Colodro-Conde¹, Paul R. H. J. Timmers², Peter K. Joshi², Grant W. Montgomery³ and Nicholas G. Martin¹

¹Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia, ²Usher Institute, University of Edinburgh, Teviot Place, Edinburgh, UK and ³Institute of Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia

Abstract

Mortality risk is known to be associated with many physiological or biochemical risk factors, and polygenic risk scores (PRSs) may offer an additional or alternative approach to risk stratification. We have compared the predictive value of common biochemical tests, PRSs and information on parental survival in a cohort of twins and their families. Common biochemical test results were available for up to 13,365 apparently healthy men and women, aged 17–93 years (mean 49.0, standard deviation [SD] 13.7) at blood collection. PRSs for longevity were available for 14,169 study participants and reported parental survival for 25,784 participants. A search for information on date and cause of death was conducted through the Australian National Death Index, with median follow-up of 11.3 years. Cox regression was used to evaluate associations with mortality from all causes, cancers, cardiovascular diseases and other causes. Linear relationships with all-cause mortality were strongest for C-reactive protein, gamma-glutamyl transferase, glucose and alkaline phosphatase, with hazard ratios (HRs) of 1.16 (95% CI [1.07, 1.24]), 1.15 (95% CI 1.04–1.21), 1.13 (95% CI [1.08, 1.19]) and 1.11 (95% CI [1.05, 1.88]) per SD difference, respectively. Significant nonlinear effects were found for urea, uric acid and butyrylcholinesterase. Lipid risk factors were not statistically significant for mortality in our cohort. Family history and PRS showed weaker but significant associations with survival, with HR in the range 1.05 to 1.09 per SD difference. In conclusion, biochemical tests currently predict long-term mortality more strongly than genetic scores based on genotyping or on reported parental survival.

Keywords: C-reactive protein; gamma-glutamyl transferase; glucose; lipids; polygenic risk scores; risk stratification

(Received 30 November 2020; accepted 7 December 2020; First Published online 29 January 2021)

Many biochemical characteristics are known to predict morbidity and mortality (Emerging Risk Factors Consortium et al., 2010; Juraschek et al., 2014; Kunutsor et al., 2014, 2015; Kuo et al., 2013; Liu et al., 2014; Long et al., 2014; Nordestgaard, 2016; Novotny & Vitek, 2003; Odden et al., 2014; Pedersen, 2016; Rader & Hovingh, 2014; Ridker et al., 2002; Stauffer et al., 2013; Wagner et al., 2015; Wu et al., 2015). A cluster of cardiovascular risk factors, including plasma lipids, are widely used for individual risk assessment. Other risk factors known to predict mortality have not been widely adopted but could supplement those currently used. In addition, longevity shows significant similarity between relatives, and whatever the cause of this similarity (see Ruby et al., 2018), a person's risk may be predictable from data on parents' age at death. Disease prediction through polygenic risk scores (PRSs) is becoming attractive as larger studies discover more loci and increase the precision of allelic risk estimates; genomewide association studies have identified single-nucleotide polymorphisms (SNPs) associated with overall mortality (Joshi et al., 2017).

Author for correspondence: John Whitfield, Email: John.Whitfield@qimrberghofer.edu.au.

Cite this article: Whitfield JB, Colodro-Conde L, Timmers PRHJ, Joshi PK, Montgomery GW, and Martin NG. (2020) Comparison of Familial, Polygenic and Biochemical Predictors of Mortality. *Twin Research and Human Genetics* 23: 307–315, <https://doi.org/10.1017/thg.2020.89>

For cardiovascular predictors, risk assessment is given value by availability of treatments that mitigate risk by lowering blood pressure or low-density lipoprotein (LDL) cholesterol (LDL-C). Absence of specific treatments may have restricted use of other predictors, but recognition of risk is a necessary step toward development of interventions. In addition to the clinical and public health value of reliable risk assessment through predictive biomarkers, development of treatments to delay age-related morbidity will require large clinical trials that may be more efficient if people at high risk, although currently well, can be selectively enrolled.

Biochemical test results that have consistently shown significant associations with mortality in the general population include lipids (Nordestgaard, 2016; Pedersen, 2016; Rader & Hovingh, 2014; Stauffer et al., 2013), inflammatory markers (Emerging Risk Factors Consortium et al., 2010; Ridker et al., 2002), liver enzymes (Kunutsor et al., 2014, 2015; Liu et al., 2014; Long et al., 2014), uric acid (Juraschek et al., 2014; Kuo et al., 2013; Odden et al., 2014; Wu et al., 2015) and bilirubin (Novotny & Vitek, 2003; Wagner et al., 2015). Despite many publications on individual risk factors, it is difficult to compare them and identify which have the strongest associations. Effective comparison of predictive biomarkers requires either combination of data across multiple studies using meta-regression (Lewington et al., 2012), or studies that obtain multiple test results on all participants.

We have taken the latter approach, using data on around 13,000 participants in twin and family studies conducted over the past 25 years. These results have been combined with follow-up through a national register of death certificate data to evaluate and compare test performance. We have evaluated linear and non-linear relationships, checked for associations specific to major cause-of-death groups and contrasted biochemical test performance against genetic indicators using parental survival data and PRSs for longevity.

Materials and Methods

Three studies on Australian adult twins and their relatives included blood collection and performance of biochemical tests. These were a twin study centered on alcohol use and alcohol use disorders (Heath *et al.*, 1997; Whitfield *et al.*, 1998); another on neuroticism and anxiety (Kirk *et al.*, 2000); and a twin-family study on alcohol and smoking (Heath *et al.*, 2011; Whitfield *et al.*, 2008). Participants in these studies were either twins who had joined a voluntary twin registry or were siblings, parents or (in a few cases) children of twins who had participated in earlier questionnaire-based twin studies. They were not selected for any disease and were healthy enough at the time of participation to be willing to take part, to respond to questionnaires or telephone interviews, and to visit a blood collection center. Approximately 45% of those with biochemical results were twins (of which 40% were from monozygotic and 60% from dizygotic pairs), 20% were parents of twins and 35% were nontwin siblings of twins. The median age at blood collection was 45.5 years, the 25th and 75th centiles were aged 37 and 59 years, and 90% were aged between 32 and 75 years. Blood collection occurred between 1993 and 1996, 1996 and 1999, and 2001 and 2005 for the three studies, and the distributions of collection dates and follow-up times are summarized in Supplementary Figure S1.

These studies conformed to the Declaration of Helsinki and had been approved by appropriate Ethics Committees, and participants gave informed consent to the data and sample collection. Death certificate matching was approved by the QIMR Berghofer Medical Research Institute Human Research Ethics Committee and the Australian Institute of Health and Welfare (AIHW) Ethics Committee.

The biochemical tests, with the number of subjects and summary statistics for each, are listed in Table 1. Testing was done using Hitachi or Roche 747, 917 or Modular P analyzers and Roche reagents (Roche Diagnostics Australia, North Ryde NSW 2113) at Royal Prince Alfred Hospital, Sydney, except that measurements of butyrylcholinesterase (BCHE) and C-reactive protein (CRP) in the 1993–96 study were performed by spectrophotometry (Valle *et al.*, 2006) and by immunonephelometry, respectively, in the laboratory of Dr Daniel O'Connor at the University of California, San Diego. Subjects were not necessarily fasting at the time of blood collection but time since the last meal was recorded and glucose results were adjusted for fasting time.

Frequency distributions for test results were plotted, and \log_{10} transformation was carried out for those with skewed distributions (alkaline phosphatase [ALP], alanine aminotransferase [ALT], aspartate aminotransferase [AST], bilirubin, CRP, gamma-glutamyl transferase [GGT], ferritin, LDL-C and total cholesterol, and triglycerides). Age-adjusted z -scores for each test were calculated separately for men and women, combined to give an age- and sex-adjusted value for each participant and used as predictor variables in survival analysis so that the hazard ratio (HR) for each test could be compared on a per-standard-deviation basis. To test for

deviation from linear relationships between test results and risk, we used the 20th, 40th, 60th and 80th centiles to define sex- and age-adjusted quintiles, and we calculated squared z -scores for each participant so that both z -score and z -score-squared could be used in survival analysis.

Body mass index (BMI, kg/m^2), habitual alcohol use (standard drinks per week) and smoking history (on a four-point scale of never/past/current < 10 cigarettes per day/current ≥ 10 cigarettes per day) were assessed using data from all studies in which these subjects participated, as described in Whitfield *et al.* (2018). Where subjects had participated in more than one study, the averages from all available time points were taken in order to obtain a better measure of lifetime exposure to these risks than a single observation.

Polygenic Risk Scores

Genotyping used Illumina arrays and imputation of additional SNPs was carried out using the Michigan imputation server, as described in the Supplementary text. Eight PRSs (PRS1 to PRS8, with p -value thresholds for SNP associations ranging from the genome-wide $p < 5 \times 10^{-8}$ to inclusion of all SNPs) were calculated using allelic association effect sizes from a genome-wide study on loci affecting longevity (Joshi *et al.*, 2017). The longevity study was based on genotyping of approximately 600,000 people who provided information on their parents' survival, with the parental survival being used as a surrogate phenotype. Data from that study were reanalyzed omitting data from participants included in the current study to avoid overlap between prediction and target samples.

Parental Survival

In previous questionnaire or interview studies, participants had answered questions about whether their mother and father were alive, and their current ages or ages at death. This information was used to generate the hazard function for each of the parents. The hazard functions were converted to Martingale residuals, which represent the deviation of each person from their expected survival, as explained in the Supplementary text. Martingale residuals for the participants' mothers and fathers, and their average (mid-parent) value, were calculated. Z -scores for these Martingale residuals were calculated using the means and standard deviations (SD s), with values below -5 set to -5 .

National Death Index Search

Names and dates of birth for participants were submitted to the Australian National Death Index (NDI; see <https://www.aihw.gov.au/about-our-data/our-data-collections/national-death-index/about-national-death-index>). Identifying information was matched against deaths occurring in Australia up to the end of October 2017. The matching algorithm used date of birth, and family and personal names weighted for frequency of names within the index. On receipt of the search results, they were ranked by matching score and checked by a person experienced with NDI data (Whitfield *et al.*, 2018).

For deaths occurring before the end of 2016, an underlying cause of death and up to 12 other conditions present were reported. Causes were missing after that time because information on date of death is received or coded earlier than the causes. Causes of death were coded by the NDI using the International Classification of Diseases, either ICD9 (up to 1996) or ICD10 (1997 onward). Only the 'underlying cause of death' was used in the cause-specific analysis. Causes were divided into three broad categories: cancers

Table 1. Statistical summary of biochemical test results for 5527 men and 7791 women with survival data and results for some or all of the listed tests

		Men								Women							
		N	Mean	SD	Centiles				N	Mean	SD	Centiles					
					20	40	60	80				20	40	60	80		
Creatinine	μmol/l	5293	100.6	18.1	86	94	102	114	7172	79.7	15.2	67	74	81	90		
Urea	mmol/l	5288	6.63	1.65	5.3	6.1	6.8	7.8	7167	5.83	1.60	4.5	5.3	6.0	7.0		
Uric acid	mmol/l	5324	0.349	0.077	0.28	0.33	0.36	0.41	7214	0.265	0.071	0.21	0.24	0.27	0.32		
Calcium	mmol/l	4576	2.450	0.150	2.35	2.42	2.48	2.56	6065	2.430	0.140	2.33	2.40	2.46	2.53		
Total protein	g/l	5225	77.0	4.8	73.0	76.0	78.0	81.0	7118	76.40	4.70	73.0	75.0	78.0	80.0		
Albumin	g/l	5287	48.7	3.0	46.0	48.0	49.0	51.0	7172	47.3	2.9	45.0	47.0	48.0	50.0		
Globulins	g/l	5216	28.3	4.1	25.0	27.0	29.0	32.0	7116	29.2	3.9	26.0	28.0	30.0	32.0		
Bilirubin	μmol/l	5256	9.90	5.50	6.0	8.0	10.0	13.0	7144	7.90	4.40	5.0	6.0	8.0	10.0		
ALP	U/l	4588	76.3	21.4	60	69	78	91	6068	72.6	24.5	53	63	74	89		
AST	U/l	5452	27.8	12.5	20	24	28	33	7579	23.0	9.3	17	20	23	27		
ALT	U/l	5452	30.0	24.5	17	22	28	38	7579	19.6	12.2	12	15	19	24		
AST/ALT ratio		5452	1.09	0.60	0.75	0.92	1.10	1.36	7578	1.34	0.53	0.95	1.17	1.36	1.64		
GGT	U/l	5453	37.0	42.0	16	22	30	46	7577	22.0	26.0	11	14	19	27		
BCHE	U/l	4410	9697	1948	8084	9208	10,151	11,325	5573	8721	1932	7086	8137	9086	10,329		
Total cholesterol	mmol/l	5525	5.69	1.06	4.80	5.40	5.90	6.60	7791	5.64	1.05	4.70	5.30	5.80	6.50		
LDL-C	mmol/l	5128	3.38	0.93	2.60	3.10	3.60	4.10	7589	3.26	0.94	2.50	3.00	3.40	4.00		
HDL-C	mmol/l	5469	1.33	0.36	1.03	1.20	1.37	1.60	7725	1.63	0.41	1.28	1.49	1.70	1.95		
Triglycerides	mmol/l	5527	2.24	1.45	1.20	1.62	2.14	3.00	7791	1.65	0.98	0.92	1.23	1.60	2.20		
Glucose	mmol/l	3562	5.36	1.90	4.34	4.77	5.14	5.83	4764	4.96	1.40	4.20	4.57	4.90	5.44		
CRP	mg/l	4401	3.25	5.83	0.8	1.3	2.2	4.1	5565	4.17	7.44	0.8	1.6	2.9	5.5		
Iron	μmol/l	4735	20.80	6.50	15.5	18.8	21.7	25.5	6543	18.50	6.60	13.1	16.5	19.5	23.5		
Transferrin	g/l	5433	2.67	0.37	2.37	2.56	2.73	2.96	7642	2.84	0.47	2.46	2.68	2.89	3.19		
Saturation	%	4733	31.3	10.5	23.0	27.9	32.5	38.7	6539	26.6	10.4	18.3	23.3	28.1	34.1		
Ferritin	mg/l	5434	251	194	102	171	249	369	7647	106	112	33	59	91	157		

Note: ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; BCHE, butyrylcholinesterase; CRP, C-reactive protein; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol.

(malignant neoplasms: ICD9 codes 140-208, ICD10 codes C00-C97); cardiovascular diseases (diseases of the circulatory system: ICD9 codes 390-459, ICD10 codes I00-I99) and other known causes. Where no cause of death was available, the date of death was used in the analysis of all-cause mortality and the case was censored (see Austin et al., 2016) at the date of death for the cause-specific analyses.

Statistical Methods

Details were submitted for NDI search on 33,593 people. Of these, biochemical results were available for 13,063, PRS for 14,169 and parental survival data (for both parents) for 25,784. Initial analyses were based on the maximum numbers available and were repeated using only those participants with biochemical data to allow comparison of effect sizes across a consistent subgroup of all participants.

Survival analysis was based on test results and the date of death or censoring, recoded to age at death or at October 21, 2017. IBM SPSS, release 22 (IBM Corp., Armonk, NY) was used for data management and estimation of means and correlations. Cox regression for survival analysis was performed using Stata (StataCorp LLC, College Station TX), with clustering by family (because of related

subjects) to generate robust standard error estimates and confidence intervals for HR. For analysis of risk by quintiles, test results were recoded using dummy variables. Analyses were repeated after exclusion of participants who had died within 2 years of blood collection to assess whether results were dominated by short-term mortality.

To compare the strength of the tests' or scores' associations with mortality, while allowing for the presence of nonlinear relationships for some tests, we calculated a predicted risk for selected tests and for each participant based on the coefficients for the z -score and z -score-squared, as determined from the initial survival analysis. These risk scores were used in a further round of Cox regression to estimate the strength of association with survival, and the β values from this analysis were used as the basis for comparison — note that this will give positive coefficients even when higher test results imply lower risk.

Results

Biochemical Tests and All-Cause Mortality

Taking each test result individually, CRP, GGT, glucose and ALP results showed significant linear associations with all-cause

mortality at the Bonferroni-corrected threshold of .0021 (for 24 comparisons; Table 2). A difference of 1 *SD* of the sample distribution for CRP was associated with a 16% difference in HR, for GGT 15%, for glucose 13% and for ALP 11%.

These estimates are estimates of linear trend, assuming the change in risk per *SD* unit is constant across the observed range. To assess more complex associations, we regressed survival on *z*-score-squared as well as *z*-score and also estimated mortality by quintile of biomarker values. Results for survival by both *z*-score and *z*-score-squared are shown in Table 2. There were significant ($p < .0021$) associations for the *z*-score-squared term for uric acid, BCHE and CRP.

Significant associations with all-cause mortality were still present for CRP, GGT, glucose and ALP when BMI, smoking status and usual alcohol intake were included as survival predictor variables in addition to sex (Supplementary Table S1), except for a slight decrease in the effect size (β) for GGT and a more substantial decrease for CRP. Results after exclusion of participants who died within 2 years of the blood collection (Supplementary Table S2) were similar to those for all participants.

HRs by age- and sex-adjusted quintiles of test results (based on comparison of the second to fifth quintiles against the first, lowest quintile) are shown in Figure 1, and more fully in Supplementary Table S3. Figure 1 shows results for CRP, GGT, glucose, ALP, uric acid, urea and albumin. Several of these show a pattern of risk that is not linear across quintiles but J- or U-shaped.

Cause-Specific Mortality

We next checked whether mortality associations were specific to one group of conditions, such as cardiovascular disease, or more general. Results are shown in Supplementary Table S4. For all-cancer mortality, significant linear or quadratic associations (at $p < .0021$) were found for albumin and CRP. For cardiovascular mortality, albumin, ALP, BCHE, creatinine, GGT, total protein and uric acid showed significant associations. For other noncancer and non-CVD conditions, CRP, glucose, iron, urea and uric acid showed significant associations. In each case, there were additional tests showing nominally significant ($p < .05$) associations.

PRSs for Mortality

PRSs were first tested for association with all-cause mortality. Results, and thresholds for inclusion of independent SNPs, are shown in Table 3. PRS1 showed the strongest and most significant association with mortality (HR per 1-*SD* difference 1.090, $p = 4.83 \times 10^{-4}$), with weaker associations for PRS2 to PRS8 (Table 3). Results obtained when only the participants with biochemical results were included are shown in Supplementary Table S5.

Results for cause-specific mortality are shown in Supplementary Table S6. For cancer deaths, there were no significant associations with PRS. For cardiovascular deaths, there were marginally significant associations for the PRS computed using more liberal *p*-value thresholds, PRS4 to PRS8. However, for deaths attributed to other causes, six out of eight scores showed significant associations, with PRS1 the strongest (HR = 1.18 per 1-*SD* difference, $p = 1.77 \times 10^{-4}$).

Parental Survival

Associations between reported parental survival and survival of the study participants, as all-cause mortality and by cause-of-death group, are shown in Table 4 and Supplementary Tables

S5 and S7. Reported parental survival was significantly associated with participant (offspring) survival, with an increase of 7% in HR for each 1-*SD* increase in mid-parent Martingale residual. The mid-parent Martingale residual was associated with risk of both cancer and cardiovascular death, with HR increasing by 6% and 21%, respectively, per 1-*SD* increase in standardized residual, but not with deaths attributed to other causes.

Biomarkers, PRSs and Parental Survival

We also considered whether the genomic risk scores, the parental survival characteristics, or both, were associated with the biochemical test results. Correlations between these are summarized in Supplementary Table S8. PRS1, based on loci known to have genomewide significant associations with survival, had significant associations with total and LDL-C, and CRP. Reported parental survival had significant correlations with albumin, bilirubin, CRP, high-density lipoprotein cholesterol (HDL-C), triglycerides and uric acid.

Discussion

Our aim was to compare prediction of mortality by widely available biochemical tests, genotype-based PRS and parental survival data. The results of this comparison are summarized in Figure 2. Our main findings are first that many of the tests associated with all-cause or grouped causes of mortality do not conform to the expectation of a linear increase or decrease in risk with increasing test results; second that lipids show no significant association with mortality in our cohort; and third that biochemical tests are currently more informative than genomic or family history data.

Lipid Risk Factors

Initially, we expected that lipids (LDL-C and HDL-C, and triglycerides) would be among the best predictors of death because of their widely used association with cardiovascular disease. In fact, lipids showed little association with all-cause mortality in this cohort (Table 2, Supplementary Table S3). Even for cardiovascular mortality (Supplementary Table S4), none of the lipid results showed significant associations. Several explanations are possible. Most participants survived to an advanced age; the estimated median life expectancy in participants with lipid results was 90 years for women and 88 for men. Many studies on lipids and cardiovascular risk have emphasized 'premature' cardiovascular mortality. Second, a decrease in the importance of lipid risk factors could occur because risk assessment and improved treatment have decreased cardiovascular death rates, including in Australia (see <https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/3303.0.55.003~1968-2017~Main%20Features~Fifty%20Years%20of%20Cardiovascular%20Mortality~2>). Use of cholesterol-lowering drugs could have affected our test results and changed the association between total or LDL-C and mortality. Either better prevention or better treatment could reduce the predictive value of lipid results relative to other markers of risk, although detection and treatment of lipid abnormalities in individual patients would remain important. Third, our lipid results were not all obtained on fasting samples. However, there is evidence (Mora *et al.*, 2019) that nonfasting lipid results give equivalent risk predictions for coronary heart disease to fasting ones.

Other Biochemical Tests

The tests that did show significant linear relationships with all-cause mortality were CRP, GGT, glucose and ALP. These point

Table 2. Cox regression for all-cause mortality using either linear (sex- and age-adjusted z-score only) or quadratic (z-score and z-score-squared) terms for test results, with familial clustering to adjust standard errors for relatedness of participants

z-scores for:	N total	N deaths	Linear				Quadratic		
			β	Robust SE	p		β	Robust SE	p
Creatinine	12,230	1369	0.0435	0.0225	.053	z	-0.0016	0.02714	.952
						z-squared	0.0243	0.0079	.0022
Urea	12,220	1364	-0.0256	0.0222	.249	z	-0.0312	0.0252	.216
						z-squared	0.0282	0.0067	2.34×10^{-5}
Uric acid	12,292	1380	0.0792	0.0251	.0016	z	0.0346	0.0259	.181
						z-squared	0.0450	0.0116	1.09×10^{-4}
Calcium	10,416	1129	-0.0151	0.0322	.639	z	-0.0202	0.0341	.554
						z-squared	-0.0063	0.0141	0.656
Total protein	12,110	11,364	-0.0031	0.0290	.914	z	0.0074	0.0271	.786
						z-squared	0.0389	0.0133	.0034
Albumin	12,225	1376	-0.0927	0.0309	.0027	z	-0.0513	0.0293	.080
						z-squared	0.0565	0.0187	.0026
Globulins	12,099	1363	0.0472	0.0269	.079	z	0.0343	0.0260	.188
						z-squared	0.0361	0.0137	.0085
Bilirubin (log)	12,166	1364	-0.0474	0.0299	.112	z	-0.0495	0.0286	.083
						z-squared	0.0280	0.0188	.137
ALP (log)	10,432	1129	0.1063	0.0309	5.82×10^{-4}	z	0.0982	0.0297	9.23×10^{-5}
						z-squared	0.0177	0.0152	.244
AST (log)	12,776	1446	-0.0029	0.0287	.919	z	-0.0039	0.0272	.887
						z-squared	0.0121	0.0081	.134
ALT (log)	12,775	1445	-0.0146	0.0317	.645	z	-0.0185	0.0284	.514
						z-squared	0.0304	0.0111	.0063
AST/ALT ratio	12,769	1444	0.0298	0.0264	.258	z	0.0117	0.0351	.738
						z-squared	0.0073	0.0076	.3342
GGT (log)	12,776	1446	0.1369	0.0275	6.70×10^{-7}	z	0.1119	0.0290	1.13×10^{-4}
						z-squared	0.0226	0.0154	.144
BCHE	11,968	1333	0.0146	0.0280	.602	z	0.0077	0.0261	.768
						z-squared	0.0391	0.0118	9.67×10^{-4}
Total cholesterol (log)	13,062	1493	0.0047	0.0264	.860	z	-0.0053	0.0251	.833
						z-squared	0.0335	0.0135	.013
LDL-C (log)	12,587	1429	0.0128	0.0267	.631	z	-0.0020	0.0265	.939
						z-squared	0.0235	0.0133	.071
HDL-C	13,063	1494	-0.0310	0.0269	.250	z	-0.0606	0.0275	.028
						z-squared	0.0451	0.0159	.0047
Triglycerides (log)	13,063	1494	0.0107	0.0275	.697	z	0.0044	0.0265	.868
						z-squared	0.0201	0.0172	.242
Glucose	8158	850	0.1258	0.0257	1.01×10^{-6}	z	0.1010	0.0379	.011
						z-squared	0.0084	0.0116	.469
CRP (log)	11,248	1233	0.1438	0.0370	1.00×10^{-4}	z	0.1449	0.0323	7.46×10^{-6}
						z-squared	0.0581	0.0151	1.23×10^{-4}
Iron	11,051	1226	-0.0599	0.0332	.071	z	-0.0699	0.0325	.031
						z-squared	0.0281	0.0164	.087
Transferrin	12,838	1470	0.0114	0.0280	.684	z	0.0055	0.0297	.852
						z-squared	0.0312	0.0173	.072

(Continued)

Table 2. (Continued)

z-scores for:	N total	N deaths	Linear			z	Quadratic		
			β	Robust SE	<i>p</i>		β	Robust SE	<i>p</i>
Transferrin saturation	11,045	1225	-0.0499	0.0334	.134	z	-0.0702	0.0319	.027
						z-squared	0.0340	0.0173	.050
Ferritin (log)	12,844	1470	0.0248	0.0258	.336	z	0.0265	0.0279	.342
						z-squared	0.0028	0.0161	.860

Note: HRs may be calculated from these results as \exp^{β} . *p* values < .0021 (0.05/24) are shown in bold type.

ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; BCHE, butyrylcholinesterase; CRP, C-reactive protein; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol. *p*-values below the Bonferroni-corrected significance threshold for 24 tests (.0021) are shown in bold.

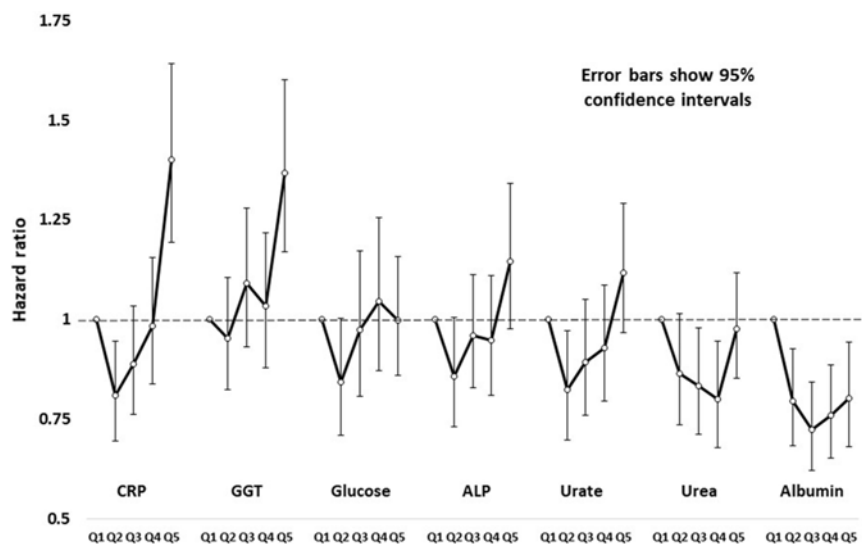


Fig. 1. Hazard ratios and robust 95% confidence intervals for all-cause mortality in study participants, by sex-specific age-adjusted quintiles of C-reactive protein (CRP), gamma-glutamyl transferase (GGT), glucose, alkaline phosphatase (ALP), urate, urea and creatinine.

to the relevance of inflammation, factors related to liver function and glycemic control.

CRP has been extensively studied, with results summarized in a meta-analysis of data from 54 prospective studies (Emerging Risk Factors Consortium *et al.*, 2010). In both men (Ridker *et al.*, 1997) and women (Ridker *et al.*, 2002), CRP and LDL-C were each risk factors for cardiovascular events, and predictive performance of CRP was better than LDL-C. Our CRP results suggest a U- or J-shaped relationship with risk (Figure 1), although the test based on assuming a linear relationship was statistically significant.

For the 'liver enzymes', many prospective studies on mortality and morbidity have been summarized and meta-analyzed, mainly with a focus on GGT and ALT (Kunutsor, Apekey, Seddoh, *et al.*, 2014; Liu *et al.*, 2014; Long *et al.*, 2014). We found significant associations with all-cause mortality for GGT and ALP but not for AST, ALT or their ratio. In this context, and because of the GGT result, we assume that the relevant source of ALP is liver rather than bone or intestine. The mechanisms linking the liver enzymes with mortality are uncertain, but at least for cardiovascular disease probably involve fatty liver (Mellinger *et al.*, 2015; VanWagner *et al.*, 2014; Wong *et al.*, 2011). In addition, GGT may have pro-oxidant effects on lipids (Franzini *et al.*, 2009; Lee & Jacobs, 2009).

It is worth noting that a significant association was found (Table 2) between all-cause mortality and glucose even though the blood samples were not collected at fasting or after a standard glucose load.

Initially, like others, we tested for linear associations with risk. However, examination of results by quintiles and tests for

quadratic associations showed that this could miss important effects. As well as CRP, renal function tests (urea, uric acid and possibly creatinine) showed a U- or J-shaped pattern of low and high results being associated with increased mortality compared to intermediate values. For urea, creatinine and uric acid renal insufficiency (at the upper end of the distributions), and poor nutritional status and muscle mass (at the lower end) could each have adverse effects. Increased mortality associated with low results for these tests could reflect ill health and short-term mortality, but our results were very similar when participants who did not survive 2 years from baseline were excluded.

Polygenic Risk Scores

We also assessed the polygenic scores for longevity, computed at various levels of significance for the SNP associations (from $p < 5 \times 10^{-8}$, to inclusion of all independent SNPs). These showed significant associations for all-cause mortality, but predictive value as measured by the HRs was less than the best of the biochemical tests, including when PRS analysis was restricted to those with biochemical results. The longevity PRS may improve with larger studies (Timmers *et al.*, 2020), although our calculation is already based on data from about 560,000 people. Also, PRS based on specific conditions may perform better than PRS for the inherently heterogeneous phenotype of all-cause mortality. It was notable that PRS1 had opposite associations with lipid risk (LDL-C, HDL-C and triglycerides) and inflammation (CRP) (Supplementary Table S7).

Table 3. Prediction of all-cause mortality from polygenic risk scores (PRSs)

	β	Robust SE	p	HR	95% CI
PRS1, $p < 5 \times 10^{-8}$	0.0861	0.0247	4.83×10^{-4}	1.090	[1.038, 1.144]
PRS2, $p < 10^{-5}$	0.0535	0.0267	0.045	1.055	[1.001, 1.112]
PRS3, $p < .001$	0.0496	0.0277	0.073	1.051	[0.995, 1.110]
PRS4, $p < .01$	0.0771	0.0275	0.0050	1.080	[1.024, 1.140]
PRS5, $p < .05$	0.0616	0.0270	0.023	1.064	[1.009, 1.121]
PRS6, $p < .1$	0.0773	0.0267	0.0039	1.080	[1.025, 1.138]
PRS7, $p < .5$	0.0712	0.0273	0.0091	1.074	[1.018, 1.133]
PRS8, $p \leq 1.0$	0.0739	0.0272	0.0065	1.077	[1.021, 1.136]

Note: HR, hazard ratio.

Standardized scores were calculated at each of the p -value thresholds shown using data from a genome-wide association study on longevity (Joshi et al., 2017). To avoid subject overlap between the data used to calculate the scores and the evaluation of the PRS predictive performance in our subjects, the allelic effect sizes were recalculated omitting data from our studies. For each PRS, the total number of participants was 14,169 and the number who died was 1645.

Table 4. Prediction of participant of all-cause mortality from parental survival data based on z -transformed Martingale residuals, winzorized at -5 , from mothers' and fathers' reported survival, and average of mother's and father's (mid-parent). There were 3740 deaths among 25,784 participants with information on both mother's and father's survival

	β	Robust SE	p	HR	95% CI
Mother	0.0520	0.0120	2.79×10^{-5}	1.053	[1.029, 1.078]
Father	0.0554	0.0142	1.84×10^{-4}	1.057	[1.028, 1.087]
Mid-parent	0.0763	0.0136	2.47×10^{-7}	1.079	[1.051, 1.109]

This will weaken the association between at least PRS1 and mortality, but this problem is inherent in the known pleiotropic (and opposite) effects on risk phenotypes at these loci.

Family History

In related questionnaire- and interview-based studies, participants had been asked whether their mother and father were living and if not, the age of each when they died. This allowed us to assess parental mortality as a predictor of mortality among participants. All-cause mortality among participants was significantly associated with their mothers' or fathers' survival, and slightly more strongly with the average of their parents' data. The parent-to-offspring transmission of risk was substantially greater for cardiovascular deaths (HR for mid-parent standardized residual 1.21) than for cancer deaths (HR 1.07) or deaths from other causes (HR 1.03) (Supplementary Table S6). This is consistent with greater heterogeneity of diseases, and presumably of heritable contributions to disease risk, for cancers and for other causes than for cardiovascular disease where atherosclerosis has a central role.

Cause-of-Death Groupings

Although the death certificate data included information on the underlying cause of death and other conditions present at the time of death, there are limitations when making associations with specific causes of death. There is a substantial literature drawing attention to inaccuracy of the causes given on death certificates (Coady

et al., 2001; Lloyd-Jones et al., 1998; Maclaine et al., 1992; Modelmog et al., 1992). Nevertheless, we did find associations between three groups of causes of death and the potential predictors.

Cancers of any type were treated as a single group. This may assume more similarity across this group than truly exists, but there were significant nonlinear associations with CRP and to a lesser extent with albumin (Supplementary Table S4). There was no significant association with the PRSs, but parental survival was significant for cancer mortality (Supplementary Tables S5 and S6). It should be noted here that we had no information on parental causes of death, so the parental survival data relate to their death from any cause.

For cardiovascular diseases, several renal or liver tests (creatinine, uric acid, total protein and albumin, ALP, BCHE and GGT) showed significant associations. The association with GGT was expected, given previous reports, but that for ALP was unexpected. The PRSs for all-cause mortality showed marginally significant associations with CVD, possibly because the loci that contribute most strongly to PRS1 have opposing effects on lipid-related and inflammation-related risk (Supplementary Table S8). The reported parental survival was significantly predictive for cardiovascular deaths.

The category of other causes of death is heterogeneous but did show associations for CRP, glucose, iron, urea and uric acid. For glucose, which is most obviously associated with diabetes and its complications, risk increased by about 15% for each 1-SD increase. CRP showed a similar 18% increase in risk for each 1-SD increase (Supplementary Table S4).

Looking at the cause-of-death grouping from another perspective, two of the best linear predictors of all-cause mortality, GGT and ALP, seem most associated with cardiovascular deaths. On the other hand, CRP and glucose show similar effect sizes across the cancers, CVD and other-cause groups (see Supplementary Table S4 and Supplementary Figure S2).

Overlap of Biochemical Results and Genetics for Mortality

Many test results in study participants were significantly correlated with the summary statistic (Martingale residuals) for survival in their parents, or with PRSs. This potentially gives novel insights about transmission of the biomarker-associated risks between generations, which will be explored in a further paper.

Strengths and Limitations

The strengths of our study are the length of follow-up and the direct comparisons made possible by having many potential predictors for the same group of subjects. However, there are larger studies, and meta-analyses covering much larger numbers of subjects, on individual risk factor/disease combinations. The age range of participants was wide, but 75% of deaths occurred after the age of 70 years and 50% over 80 years. Therefore, our results mostly reflect mortality associations in older people. Also, many potential associations have been examined and some of the results (e.g. subdivision by major disease groupings) should be regarded as hypothesis-generating and in need of replication. Maturation of large studies such as the UK Biobank (<http://www.ukbiobank.ac.uk/>) will facilitate this. Observational data alone cannot show causation, but examination of genetic effects on these biomarkers and on mortality using the Mendelian randomization approach should clarify causal relationships.

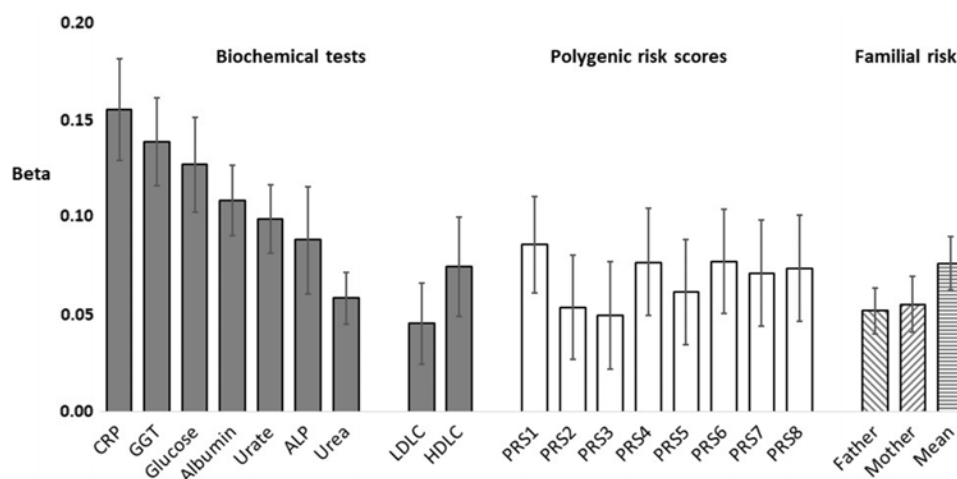


Fig. 2. Comparison of estimated effects (β -coefficients and standard errors from Cox regression) on all-cause mortality for selected biochemical tests, polygenic risk scores and parental survival data. Risk estimates for the biochemical tests were recalculated for this figure using both linear and quadratic coefficients for predicted risk as listed in Table 2. PRS1 to PRS8; polygenic risk scores; father, mother and mean; parental Martingale residuals (based on their reported survival) for mothers, fathers and mean (mid-parent) value

Conclusions

Comparison between tests shows that inflammation (CRP), liver function (GGT, ALP) and glycaemic control (glucose) show the strongest associations with mortality, but indicators of adverse nutrition (low urea) also affect survival. The widely used lipid risk factors were less predictive than expected. Polygenic scores for mortality show promise and may improve.

Acknowledgments. We are grateful to the participants who provided information and blood samples on which this report is based; to the staff of QIMR Berghofer Medical Research Institute, Brisbane for liaison with participants and data collection and management and particularly to Richard Parker for organizing the NDI search and Judith Symmons for checking the matched results from the search; and to the staff of the Biochemistry Department, Royal Prince Alfred Hospital, Sydney who performed the biochemical tests. Many of the twin participants were originally contacted through the Australian Twin Registry.

Financial Support. Funding for the performance of the biochemical tests came from the US National Institutes of Health (AA14041), which also funded aspects of participant recruitment and sample collection (AA07535, AA07728, AA11998, AA13320, AA13321, AA17688, DA012854, DA019951). The National Death Index search was made possible by a donation from Mr George Landers. Funders had no role in study design, in the collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit the article for publication.

Conflict of Interest. None.

Data Statement. Because of undertakings given to the participants as part of the informed consent process, raw data on which our results are based cannot be made freely available. However, it may be possible for interested researchers to make use of the data during a prearranged visit to QIMR Berghofer Medical Research Institute.

Supplementary Material. To view supplementary material for this article, please visit <https://doi.org/10.1017/thg.2020.89>.

References

- Austin, P. C., Lee, D. S., & Fine, J. P. (2016). Introduction to the analysis of survival data in the presence of competing risks. *Circulation*, *133*, 601–609.
- Coady, S. A., Sorlie, P. D., Cooper, L. S., Folsom, A. R., Rosamond, W. D., & Conwill, D. E. (2001). Validation of death certificate diagnosis for coronary heart disease: The Atherosclerosis Risk in Communities (ARIC) Study. *Journal of Clinical Epidemiology*, *54*, 40–50.
- Emerging Risk Factors Consortium, Kaptoge, S., Di Angelantonio, E., Lowe, G., Pepys, M. B., Thompson, S. G., & Danesh, J. (2010). C-reactive protein

concentration and risk of coronary heart disease, stroke, and mortality: An individual participant meta-analysis. *Lancet*, *375*, 132–140.

- Franzini, M., Corti, A., Martinelli, B., Del Corso, A., Emdin, M., Parenti, G. F., & Paolicchi, A. (2009). Gamma-Glutamyltransferase activity in human atherosclerotic plaques—biochemical similarities with the circulating enzyme. *Atherosclerosis*, *202*, 119–127.
- Heath, A. C., Bucholz, K. K., Madden, P. A., Dinwiddie, S. H., Slutske, W. S., Bierut, L. J., & Martin, N. G. (1997). Genetic and environmental contributions to alcohol dependence risk in a national twin sample: Consistency of findings in women and men. *Psychological Medicine*, *27*, 1381–1396.
- Heath, A. C., Whitfield, J. B., Martin, N. G., Pergadia, M. L., Goate, A. M., Lind, P. A., & Montgomery, G. W. (2011). A quantitative-trait genome-wide association study of alcoholism risk in the community: Findings and implications. *Biological Psychiatry*, *70*, 513–518.
- Joshi, P. K., Pirastu, N., Kentistou, K. A., Fischer, K., Hofer, E., Schraut, K. E., & Wilson, J. F. (2017). Genome-Wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. *Nature Communications*, *8*, 910.
- Juraschek, S. P., Tunstall-Pedoe, H., & Woodward, M. (2014). Serum uric acid and the risk of mortality during 23 years follow-up in the Scottish Heart Health Extended Cohort Study. *Atherosclerosis*, *233*, 623–629.
- Kirk, K. M., Birley, A. J., Statham, D. J., Haddon, B., Lake, R. I., Andrews, J. G., & Martin, N. G. (2000). Anxiety and depression in twin and sib pairs extremely discordant and concordant for neuroticism: Prodigium to a linkage study. *Twin Research*, *3*, 299–309.
- Kunutsor, S. K., Apekey, T. A., & Khan, H. (2014). Liver enzymes and risk of cardiovascular disease in the general population: A meta-analysis of prospective cohort studies. *Atherosclerosis*, *236*, 7–17.
- Kunutsor, S. K., Apekey, T. A., Seddoh, D., & Walley, J. (2014). Liver enzymes and risk of all-cause mortality in general populations: A systematic review and meta-analysis. *International Journal of Epidemiology*, *43*, 187–201.
- Kunutsor, S. K., Apekey, T. A., Van Hemelrijck, M., Calori, G., & Perseghin, G. (2015). Gamma glutamyltransferase, alanine aminotransferase and risk of cancer: Systematic review and meta-analysis. *International Journal of Cancer*, *136*, 1162–1170.
- Kuo, C. F., See, L. C., Yu, K. H., Chou, I. J., Chiou, M. J., & Luo, S. F. (2013). Significance of serum uric acid levels on the risk of all-cause and cardiovascular mortality. *Rheumatology (Oxford)*, *52*, 127–134.
- Lee, D. H., & Jacobs, D. R., Jr. (2009). Serum gamma-glutamyltransferase: New insights about an old enzyme. *Journal of Epidemiology and Community Health*, *63*, 884–886.
- Lewington, S., Bragg, F., & Clarke, R. (2012). A review on metaanalysis of biomarkers: Promises and pitfalls. *Clinical Chemistry*, *58*, 1192–1204.
- Liu, Z., Ning, H., Que, S., Wang, L., Qin, X., & Peng, T. (2014). Complex association between alanine aminotransferase activity and mortality in general population: A systematic review and meta-analysis of prospective studies. *PLoS One*, *9*, e91410.

- Lloyd-Jones, D. M., Martin, D. O., Larson, M. G., & Levy, D. (1998). Accuracy of death certificates for coding coronary heart disease as the cause of death. *Annals of Internal Medicine*, *129*, 1020–1026.
- Long, Y., Zeng, F., Shi, J., Tian, H., & Chen, T. (2014). Gamma-Glutamyltransferase predicts increased risk of mortality: A systematic review and meta-analysis of prospective observational studies. *Free Radical Research*, *48*, 716–728.
- Maclaine, G. D., Macarthur, E. B., & Heathcote, C. R. (1992). A comparison of death certificates and autopsies in the Australian Capital Territory. *Medical Journal of Australia*, *156*, 462–463, 466–468.
- Mellinger, J. L., Pencina, K. M., Massaro, J. M., Hoffmann, U., Seshadri, S., Fox, C. S., & Speliotes, E. K. (2015). Hepatic steatosis and cardiovascular disease outcomes: An analysis of the Framingham Heart Study. *Journal of Hepatology*, *63*, 470–476.
- Modelmog, D., Rahlenbeck, S., & Trichopoulos, D. (1992). Accuracy of death certificates: A population-based, complete-coverage, one-year autopsy study in East Germany. *Cancer Causes Control*, *3*, 541–546.
- Mora, S., Chang, C. L., Moorthy, M. V., & Sever, P. S. (2019). Association of nonfasting vs fasting lipid levels with risk of major coronary events in the Anglo-Scandinavian Cardiac Outcomes Trial-Lipid Lowering Arm. *JAMA Internal Medicine*. *179*, 898–905.
- Nordestgaard, B. G. (2016). Triglyceride-Rich lipoproteins and atherosclerotic cardiovascular disease: New insights from epidemiology, genetics, and biology. *Circulation Research*, *118*, 547–563.
- Novotny, L., & Vitek, L. (2003). Inverse relationship between serum bilirubin and atherosclerosis in men: A meta-analysis of published studies. *Experimental Biology and Medicine*, *228*, 568–571.
- Odden, M. C., Amadu, A. R., Smit, E., Lo, L., & Peralta, C. A. (2014). Uric acid levels, kidney function, and cardiovascular mortality in US adults: National Health and Nutrition Examination Survey (NHANES) 1988–1994 and 1999–2002. *American Journal of Kidney Diseases*, *64*, 550–557.
- Pedersen, T. R. (2016). The success story of LDL cholesterol lowering. *Circulation Research*, *118*, 721–731.
- Rader, D. J., & Hovingh, G. K. (2014). HDL and cardiovascular disease. *Lancet*, *384*, 618–625.
- Ridker, P. M., Cushman, M., Stampfer, M. J., Tracy, R. P., & Hennekens, C. H. (1997). Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *New England Journal of Medicine*, *336*, 973–979.
- Ridker, P. M., Rifai, N., Rose, L., Buring, J. E., & Cook, N. R. (2002). Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. *New England Journal of Medicine*, *347*, 1557–1565.
- Ruby, J. G., Wright, K. M., Rand, K. A., Kermany, A., Noto, K., Curtis, D., & Ball, C. (2018). Estimates of the heritability of human longevity are substantially inflated due to assortative mating. *Genetics*, *210*, 1109–1124.
- Stauffer, M. E., Weisenfluh, L., & Morrison, A. (2013). Association between triglycerides and cardiovascular events in primary populations: A meta-regression analysis and synthesis of evidence. *Vascular Health and Risk Management*, *9*, 671–680.
- Timmers, P., Wilson, J. F., Joshi, P. K., & Deelen, J. (2020). Multivariate genomic scan implicates novel loci and haem metabolism in human ageing. *Nature Communications*, *11*, 3570.
- Valle, A., O'Connor, D. T., Taylor, P., Zhu, G., Montgomery, G. W., Slagboom, P. E., & Whitfield, J. B. (2006). Butyrylcholinesterase: Association with the metabolic syndrome and identification of 2 gene loci affecting activity. *Clinical Chemistry*, *52*, 1014–1020.
- VanWagner, L. B., Ning, H., Lewis, C. E., Shay, C. M., Wilkins, J., Carr, J. J., & Carnethon, M. R. (2014). Associations between nonalcoholic fatty liver disease and subclinical atherosclerosis in middle-aged adults: The Coronary Artery Risk Development in Young Adults Study. *Atherosclerosis*, *235*, 599–605.
- Wagner, K. H., Wallner, M., Molzer, C., Gazzin, S., Bulmer, A. C., Tiribelli, C., & Vitek, L. (2015). Looking to the horizon: The role of bilirubin in the development and prevention of age-related chronic diseases. *Clinical Science*, *129*, 1–25.
- Whitfield, J. B., Dy, V., Madden, P. A., Heath, A. C., Martin, N. G., & Montgomery, G. W. (2008). Measuring carbohydrate-deficient transferrin by direct immunoassay: Factors affecting diagnostic sensitivity for excessive alcohol intake. *Clinical Chemistry*, *54*, 1158–1165.
- Whitfield, J. B., Fletcher, L. M., Murphy, T. L., Powell, L. W., Halliday, J., Heath, A. C., & Martin, N. G. (1998). Smoking, obesity, and hypertension alter the dose-response curve and test sensitivity of carbohydrate-deficient transferrin as a marker of alcohol intake. *Clinical Chemistry*, *44*, 2480–2489.
- Whitfield, J. B., Heath, A. C., Madden, P. A. F., Landers, J. G., & Martin, N. G. (2018). Effects of high alcohol intake, alcohol-related symptoms and smoking on mortality. *Addiction*, *113*, 158–166.
- Wong, V. W., Wong, G. L., Yip, G. W., Lo, A. O., Limquiaco, J., Chu, W. C., & Chan, H. L. (2011). Coronary artery disease and cardiovascular outcomes in patients with non-alcoholic fatty liver disease. *Gut*, *60*, 1721–1727.
- Wu, C. Y., Hu, H. Y., Chou, Y. J., Huang, N., Chou, Y. C., Lee, M. S., & Li, C. P. (2015). High serum uric acid levels are associated with all-cause and cardiovascular, but not cancer, mortality in elderly adults. *Journal of the American Geriatrics Society*, *63*, 1829–1836.