BMJ Open

Primary care use of laboratory tests in Northern Ireland's Western Health and Social Care Trust: a cross-sectional study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-026647
Article Type:	Research
Date Submitted by the Author:	12-Sep-2018
Complete List of Authors:	Bucholc, Magda; University of Ulster - Magee Campus, School of Computing, Engineering and Intelligent Systems OKane, Maurice; Altnagelvin Hospital, Clinical Chemistry Mullan, Ciaran; Health and Social Care Board Western Local Commissioning Group Ashe, Siobhan; Altnagelvin Area Hospital Maguire, Liam; University of Ulster - Magee Campus, School of Computing, Engineering and Intelligent Systems.
Keywords:	Pathology < BASIC SCIENCES, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PATHOLOGY, Chemical pathology < PATHOLOGY, PRIMARY CARE

Primary care use of laboratory tests in Northern Ireland's Western Health and Social Care Trust: a cross-sectional study.

Magda Bucholc[1], Maurice J O'Kane[2], Ciaran Mullan[3], Siobhan Ashe[2], Liam Maguire[1]

[1] Intelligent Systems Research Centre, University of Ulster, Magee Campus, Londonderry BT48 7JL, Northern Ireland, UK

[2] Altnagelvin Area Hospital, Western Health and Social Care Trust, Glenshane Road, Londonderry BT47 6SB, Northern Ireland, UK

[3] Western Local Commissioning Group, Gransha Park House, Clooney Road, Londonderry BT 47 6FN, Northern Ireland, UK

Correspondence to

Dr Magda Bucholc; m.bucholc@ulster.ac.uk

Abstract

Objectives To describe the laboratory test ordering patterns by general practitioners (GPs) in Northern Ireland Western Health and Social Care Trust (WHSCT) and establish demographic and socio-economic determinants of test requesting.

Design Cross-sectional study.

Setting Western Health and Social Care Trust, Northern Ireland

Participants 55 WHSCT general practices requesting laboratory tests in the period from 1 April 2011 to 31 March 2016

Outcomes To identify the temporal patterns of laboratory test ordering behaviour for 8 commonly requested clinical biochemistry tests/test groups in WHSCT. To analyse the extent of variations in laboratory test requests by GPs and to determine whether these variations can be accounted for by clinical outcomes or geographical, demographic, and socioeconomic characteristics.

Results We identified substantial changes in the median number of request rates over five consecutive years of the study period as well as a large variation of adjusted test request rates for individual tests (lowest for electrolyte profiles, liver profiles, and HbA_{1c} and highest for immunoglobulins). There was no significant relationship between ordering activity and either demographic (age and gender) and socioeconomic factors (deprivation) or Quality and Outcome Framework (QOF) scores. We found that practice setting accounted for some of the between-practice variation in test requesting. Rural practices were characterized by both higher between practice variability and median number of order tests than urban practices at all time points.

Conclusions A large between-practice variation in GP laboratory test requesting, unrelated to demographic and socioeconomic indicators of the practices or crude clinical outcome indicators, most likely reflects differences in the clinical practice of individuals, potentially amenable to change through clinical interventions.

Strengths and limitations of this study

- The study provides a comprehensive analysis of temporal changes in laboratory test utilization patterns and establishes the extent of variability in test requesting activity across general practices in Northern Ireland's Western Health and Social Care Trust.
- The substantial variation in test ordering, not related to demographic and socioeconomic characteristics of practices, practice location or clinical outcome indicators, may reflect inappropriate laboratory test utilization and hence, suggest a potential for more efficient demand management of laboratory services.
- Given a cohort of general practices within one catchment area, our results provide evidence of differences in behaviour of individual GPs when managing patients with similar clinical symptoms.
- Failure to collect and cross-tabulate data on characteristics of general practitioners (GPs), such as GP's age, years of experience, medical training was a study limitation and a missed opportunity in assessing the influence of practitioner factors on the variation in test ordering behaviour.

Introduction

Despite the important role of laboratory testing in the diagnosis and monitoring of disease, there is concern about the increasing number of requested tests and in particular, large differences in laboratory utilization between clinical teams.[1] In the UK, laboratory test orders grew by approximately 5% per year in recent years and inappropriate test requests are considered to be an important cause for this increase[2,3]. Although pathology expenditures account for only 5–6% of the UK total health budget, they are viewed as a potential source of savings, most likely because the costs can be easily identified and measured.[4] According to the Department of Health, the rationalization of pathology services, including demand management of laboratory tests and elimination of unnecessary requesting, could produce savings of at least £500 million.[5]

Unnecessary testing is not only wasteful of resources but impacts on patients directly through the requirement for venepuncture and the follow up of minor (and possibly insignificant) abnormalities which may cause patient anxiety. On the other hand, inappropriate under requesting may cause harm through failure to diagnose or manage disease optimally. Several studies suggested that unnecessary and inappropriate utilization of laboratory services is closely linked to inter-practitioner variability in test requesting.[6-8] Despite the increased availability of clinical management guidelines promoting harmonization of the use of laboratory tests, there is still substantial variation in test utilization among general practitioners.[7,9] These differences appear to be unrelated to demographic characteristics of patient populations, socio-economic status of GP practices, disease prevalence or clinical outcome indicators.[6,7,10,11] Even if some of these variables have been shown to have an effect on test ordering patterns, the variation in requesting rates is so large that it can only be explained by differences

in attitudes towards the use of laboratory tests of individual practitioners.[10] Accordingly, factors such as confidence in clinical judgement, clinical experience, an attitude about clinical practice guidelines, a lack of knowledge regarding the correct use of tests and fear of litigation have been identified as potential sources of practice variation.[12-15]

Since unwarranted variation can lead to suboptimal clinical outcomes,[16,17] identification of determinants contributing to differences in test requesting can provide useful information for optimising utilization of laboratory services. The aim of our study was to establish the extent of variability in test requesting and characterise temporal changes in test ordering patterns across general practices within the catchment area of the Northern Ireland (NI) Western Health and Social Care Trust (WHSCT) for a range of commonly requested clinical biochemistry tests/test groups. In addition, we investigated potential determinants of inter-doctor variability in the use of laboratory tests including geographical, demographic, and socioeconomic factors as well as Quality and Outcome Framework (QOF) scores.

Materials and Methods

Study design and data sources

We conducted a cross-sectional study of laboratory test ordering activity across general practices in the WHSCT in the period from 1 April 2011 to 31 March 2016. The data on the use of laboratory tests were obtained from a HSC Business Object XI clinical information system. We investigated requesting rates for 8 clinical biochemistry tests/test groups including electrolyte profile, thyroid profile (FT4 and TSH), liver profile, lipid profile, urine albumin/creatinine (ACR) ratio, glycosylated haemoglobin (HbA_{1c}), prostate-specific antigen (PSA), and immunoglobulins. The number of laboratory tests requested in each general practice was normalized by the number of

registered patients and expressed as requests per 1000 patients, with the exception of H_bA_{1c} and ACR for which ordering rates were expressed as tests per patient with diabetes and PSA rates, calculated per 1000 male patients. GP practice list size data (including the number of patients with diabetes) and patient demographic data was extracted from the Business Services Organisation (BSO) Family Practitioner Service Information and Registration Unit system. The rural-urban distribution of GP practices was based on the data from the Census Office of the Northern Ireland Statistics and Research Agency (NISRA).[18] Socioeconomic characteristics of GP practices were determined using the NISRA Neighbourhood Information System (NINIS).[19] QOF scores for individual practices were extracted from the website of Northern Ireland Department of Health.[20]

Participants and setting

Data on laboratory tests requested from 55 general practices within the WHSCT were collected from the laboratory databases of Clinical Chemistry departments of the Altnagelvin Area Hospital, Tyrone County Hospital, and the Erne Hospital (subsequently the South West Acute Hospital). WHSCT provides health and social care services throughout the west of Northern Ireland, across the council areas of Derry City and Strabane District Council, Limavady in the Causeway Coast and Glens Borough Council, and Fermanagh and Omagh District Council.

Inclusion criteria

We examined only laboratory test requests from 55 separate primary care medical practices within the catchment area of WHSCT that remained open throughout the study period i.e. during five consecutive years: Apr 2011 – Mar 2012, Apr 2012 – Mar 2013, Apr 2013 – Mar 2014, Apr 2014 – Mar 2015, and Apr 2015 – Mar 2016, were

examined. To ensure the completeness and consistency of the data, test orders from WHSCT practices that closed or merged during the period of investigation were not taken into account.

Variables and characteristics

Our analysis was limited to 8 frequently requested clinical biochemistry tests/test groups including electrolyte profile, thyroid profile, liver profile, lipid profile, ACR, HbA_{1c}, PSA, and immunoglobulins. All considered tests/profiles (with the exception of immunoglobulins and ACR) were listed on a paper pathology request form and ordered by ticking a box adjacent to the test name; requests for ACR and immunoglobulins were made by writing the test name in an open text field.

Rates of test requests were analysed in the context of geographical and socioeconomic characteristics of practices, patient demographics (age and gender), and clinical outcome indicators. Note that we were not able to obtain a consolidated data on both gender and age of patients registered in individual GP practices.

The practice setting (rural v. urban area) was determined using the urban-rural classification of the Department for Environment, Food and Rural Affairs (DEFRA)[21] while the size of settlements was obtained from the NISRA Census Office.[18] A GP practice was defined as rural if its physical address was situated in a settlement of less than 10,000. Accordingly, we classified 24 practices as rural and 31 as urban. Furthermore, GP practices were categorised based on the Northern Ireland Multiple Deprivation Measure (MDM) identified for individual Super Output Areas (SOAs).[19] The MDM comprises a weighted combination of 7 component measures (Income Deprivation; Employment Deprivation; Health Deprivation and Disability; Education, Skills and Training Deprivation; Proximity to Services; Living Environment; and Crime

and Disorder) and ranges from 1 (most deprived SOAs) to 890 (least deprived SOAs). The Health Deprivation & Disability rank (one of the MDM domains)[19] was analysed separately as a potential determinant of inter-doctor variability in the use of laboratory tests.

Given a detailed breakdown of the age of patients on the individual practice lists, we examined the relationship between the percentage of patients over age 65 and ordering rates for electrolyte, thyroid, liver, and lipid profiles. In addition, we investigated the link between gender and requesting activity for PSA (for males) and thyroid profiles (for females) and evaluated test requesting patterns for diabetes mellitus and their relationship to QOF clinical indicator scores in diabetes.

Outcome measures

Our outcome of interest was to identify the presence and extent of variations in primary care laboratory tests ordering and to evaluate temporal changes in both the standardised number of test requests and between-practice variability in requesting. In addition, we studied demographic, socio-economic, geographical, and clinical factors that may explain this variation.

Statistical analysis

Inter-practice variability in test requests was assessed by calculating the variance (σ^2). Furthermore, we computed 'variability index' (Var_i) defined as the top decile divided by the bottom decile of standardized test request rates.[10] Var_i is a dimensionless measure of dispersion allowing us to compare the amount of variation in request rates of individual tests despite their differences in scale. The Shapiro-Wilk test of normality was used to determine if the distribution of test ordering data deviated from a normal distribution.[22] Since the distribution of laboratory test request rates was found to be

non-normal, the non-parametric statistics were implemented to perform further analysis. Mann Whitney U (MWU) test was employed to compare distributions of laboratory test rates between GP practices located in rural and urban areas.[23] The homogeneity of variances for requesting activity in rural and urban GP practices was assessed with the Fligner-Killeen (FK) test.[24] Significance of temporal changes in median and variability in test requesting rates was examined with the Mann–Kendall (MK) test.[25] In all conducted tests, a $p < 0.05$ was considered significant. Kendall rank correlation coefficient (τ)[26] was calculated to test the relationship between adjusted requesting rates and 1) Multiple Deprivation Measure; 2) Health Deprivation & Disability Deprivation rank; 3) proportions of patients over age 65; 4) distribution of patient's gender; and 5) QOF clinical indicator scores. The Kendall coefficient ($-1 \leq \tau \leq 1$) measures the strength of a correlation between two variables and assesses the degree of overall correspondence of variables' ranking.[26]

Patient and public involvement

No patients or general practitioners were involved in defining the research question or outcome measures nor were they engaged in the design or implementation this study. The study was approved by the Western Health and Social Care Trust (WHSCT). All identifiable information used for the purpose of the study was anonymised and not traceable to individual patients or general practitioners.

Results

Test ordering patterns

We analysed the laboratory test request rates of 55 general practices within the catchment area of the WHSCT comprising a total of 316 382 (2011-12), 316 688

(2012-13), 318 057 (2013-14), 319 383 (2014-2015), and 326 429 (2015-16) patients. Figure 1 shows the total number of 8 considered clinical biochemistry tests/test groups ordered during the study period of five consecutive years. The total number of ordered tests was 523 111 (2011-12), 531 849 (2012-13), 531 583 (2013-14), 525 146 (2014-2015), and 542 118 (2015-16) with electrolyte profiles found to be the most frequently ordered tests, making up approximately 30% of all requests in each year.

We observed substantial between-practice variability for all studied tests (table 1). In the considered period, the Var_i was lowest for electrolyte profiles (2.0-2.3), liver profiles (2.2-2.5), and HbA_{1c} (2.2-2.9) and highest for immunoglobulins reaching the value of 69.8 in 2012-13. The large variability was caused by several GP practices with abnormally high/low ordering rates of laboratory tests (figure 2). For the majority of tests, the median of test request rates was found to be generally lower than the mean, suggesting that the distribution of ordered test was primarily affected by the presence of GP practices with extremely high numbers of standardized requests.

Changes in the number and variation of test requests over time

Over the period of investigation, we observed an increase in the median standardized number of electrolyte profiles, immunoglobulins, PSA, and HbA_{1c} (figure 3); however, the Mann–Kendall (MK) test indicated the statistically significant upward trend only for requesting rates of PSA (p_{MK} = 0.03) and HbA_{1c} (p_{MK} = 0.03). For lipid profiles, thyroid profiles, liver profiles, and ACR, we reported the overall decline of the median number of test request rates. Yet, the downward trend was found statistically significant only for lipid profiles (p_{MK} = 0.03).

Table 1. Overview of general statistics

	EP	Lipids	PSA	Thyroid	HbA1c	Liver	ACR	Immunos
2011-2012								
Mean, 95%CI	511.4, (477.6, 545.2)	282.7, (250.2, 315.3)	69.4, (56.7, 82.0)	312.8, (278.3, 347.4)	1.86, (1.74, 1.98)	442.4, (408.0, 476.8)	1.91, (1.73, 2.10)	4.9, (3.9, 6.0)
Median, 95%CI	520.5, (476.6, 536.2)	268.9, (243.8, 285.5)	53.2, (46.6, 61.6)	296.7, (258.1, 347.7)	1.75, (1.66, 1.84)	435.7, (390.7, 469.2)	1.86, (1.60, 1.98)	4.1, (2.6, 5.2)
Range	267 - 1002	101.8 - 954.5	19.6 - 279.3	121 - 969.4	1.1 - 3.1	254 - 989.1	0.6 - 3.8	0.3 - 16.8
Variance	15660.2	14519.1	2193.8	16320.0	0.2	16187.4	0.5	15.8
Var_i	2.3	3.4	7.0	3.3	2.2	2.5	3.4	18.3
2012-2013								
Mean, 95%CI	516.1, (483.7, 548.4)	274.1, (241.5, 306.6)	79.2, (62.3, 96.0)	315.9, (280.4, 351.4)	2.04, (1.90, 2.19)	443.8, (410.7, 476.9)	1.90, (1.70, 2.10)	5.9, (4.5 7.3)
Median	491.7, (482.0, 530.6)	257.4, (238.9 280.6)	59.2, (48.2, 68.3)	291.7, (261.0, 352.0)	1.98, (1.80, 2.14)	432.3, (406.2, 462.6)	1.70, (1.60, 1.95)	4.5, (3.4, 5.6)
Range	283.6 - 1041.8	105.8 - 986.3	19.9 - 396.1	115.5 - 995.2	1.1 - 3.4	258.7 - 1016.4	0.5 - 4.8	0 - 24
Variance	14299.4	14494.8	3896.4	17230.2	0.3	14997.6	0.5	28.0
Var_i	2.2	3.4	7.6	3.3	2.4	2.4	3.7	69.8
2013-2014								
Mean, 95%CI	511.1, (480.4, 541.7)	258.7, (227.7, 289.7)	79.6, (60.2, 99.0)	295.4, (262.2, 328.6)	2.32, (2.11, 2.52)	427.6, (395.9, 459.3)	1.92, (1.73, 2.12)	6.5, (4.9, 8.2)
Median	508.8, (482.0, 530.6)	239.3, (213.7, 268.2)	59.5, (54.0, 70.7)	264.3, (244.1, 289.8)	2.10, (1.91, 2.38)	408.3, (379.6, 444.4)	1.81, (1.62, 2.03)	4.4, (3.0, 6.1)
Range	325.1 - 1014.9	111 - 950.7	23.1 - 527.6	141.5 - 962.2	1.3 - 4.6	267.1 - 987.8	0.7 - 4.7	0.5 - 24.8
Variance	12831.5	13134.3	5153.6	15099.2	0.6	13739.4	0.5	36.2
Var_i	2.0	3.1	7.2	3.0	2.8	2.3	3.3	19.2
2014-2015								
Mean, 95%CI	515.0, (484.5, 545.4)	238.4, (206.6, 270.1)	74.7, (62.1, 87.4)	290.7, (258.4, 323.1)	2.60, (2.36, 2.85)	432.8, (401.6, 463.9)	1.45, (1.27, 1.64)	8.2, (5.9, 10.4)
Median	508.7, (482.0, 522.2)	224.0, (202.4, 238.1)	63.9, (55.4, 69.0)	260.4, (244.1, 286.9)	2.33, (2.23, 2.55)	422.4, (388.0, 450.7)	1.31, (1.14, 1.60)	5.9, (4.1, 8.3)
Range	274.9 - 976.9	82.8 - 939.5	17.1 - 274	152.7 - 951.7	1.5 - 5.9	260.7 - 968	0.2 - 4	0 - 46.6
Variance	12696.9	13814.0	2192.6	14324.6	0.8	13251.9	0.5	67.3
Var_i	2.1	3.5	6.3	3.0	2.9	2.3	4.7	33.6
2015-2016								
Mean, 95%CI	526.3, (495.5, 557.2)	232.5, (202.6, 262.5)	82.9, (68.2, 97.6)	295.0, (265.8, 324.3)	3.01, (2.72, 3.31)	435.2, (405.1, 465.3)	1.21, (0.98, 1.44)	8.5, (6.4, 10.6)
Median	527.5, (462.3, 553.7)	216.6, (189.4, 235.5)	68.9, (58.6, 76.2)	268.4, (249.4, 302.1)	2.81, (2.55, 2.98)	426.3, (402.6, 439.8)	1.01, (0.85, 1.25)	5.6, (3.7, 9.5)
Range	316 - 925.3	96.9 - 846.5	26.4 - 296.9	142.5 - 857.2	1.7 - 8.1	274.6 - 898.3	0.1 - 5	0 - 33.9
Variance	13048.3	12303.9	2961.0	11708.1	1.2	12378.9	0.7	58.9
Var_i	2.0	3.7	6.5	2.8	2.9	2.2	9.4	24.5

The variance for electrolyte profiles, lipid profiles, thyroid profiles, and liver profiles requests fell by 16.8%, 15.3%, 28.3%, and 23.5% respectively (figure 3). The increase in variance was observed for immunoglobulins (272%), ACR (40%), and HbA_{1c} (500%). Nonetheless, only the increase in variance of HbA_{1c} requests (p_{MK} = 0.03) and decrease in variance of standardized number of liver profiles (p_{MK} = 0.03) were found significant at the 95% confidence level. For PSA test request rates, we observed fluctuations in variance with no clear trend.

Test ordering variation between general practices and explanatory factors

The relationship between test requesting rates and potential explanatory factors was established based on the information on the number of ordered tests, patient demographics, and Quality and Outcomes Framework indicator scores obtained for the period 1 Apr 2015 – 31 Mar 2016.

Demographic characteristics of patient population

Proportion of the oldest age category of patients (> 75) constituted a relatively small group in each GP practice (mean = 6.1%, 95%CI = (5.7%,6.5%)). Hence, we combined the 65-74 and > 75 age categories to create a more meaningful group of patients in the older age band. We examined four tests for which differences in proportions of older patients were expected to be reflected in test requesting i.e. electrolyte profiles, liver profiles, lipid profiles, and thyroid profiles. Since we were unable to obtain a consolidated data on both gender and age of patients, we did not assess the relationship between the PSA requesting rates and the category of males aged 65 and over. We found a very weak correlation between the adjusted request rates of selected laboratory tests and the percentage distribution of patients of age over 65, with τ ranging from 0.08 for thyroid profiles to 0.23 for lipid profiles (supplementary table S1).

Since previous studies reported on a higher rate of testing of thyroid hormones in females,[27] we looked at the strength of a relationship between the percentage of females in individual GP practices and requesting rates for thyroid profiles. We found a weak, in fact negative, association between these two characteristics, with $\tau = -0.2$ (supplementary figure S2).

In addition, we examined the effect of gender distribution on the standardized number of PSA tests, with the PSA ordering rates expected to be higher for GP practices with higher percentage of males (supplementary figure S2). The coefficient $\tau = 0.2$ implied a weak degree of correlation between these two variables. Note that we acknowledge the fact that a combined effect of sex and age distributions might have had a more significant effect on PSA requesting activity. However, we were unable to extract such data.

Practice setting

Figure 4 shows a temporal median-variance relationship of the standardized number of laboratory test requests for rural and urban areas. Large differences were identified for PSA, lipid profiles, thyroid profiles, and liver profiles. The median number of tests ordered annually by practices located in rural areas was higher by approximately 27-37% for PSA, 14-30% for lipid profiles, 14-38% for thyroid profiles, and 8-23% for liver profiles. For ACR and immunoglobulins the median requesting rates were lower in rural areas by 1-27% and 18-57% respectively. Across five consecutive time periods, Mann Whitney U (MWU) test showed significant differences between 'rural' and 'urban' distributions of laboratory test rates for lipid profiles (p_{MWU} ranging from 0.00004 to 0.03) and PSA (p_{MWU} between 0.003 and 0.03). The significant rural-urban differences in requesting activity of thyroid and liver profies ($p_{MWU} < 0.05$) were observed only in the first two years of the study (table 2).

Rural practices had significantly higher variability in test requesting than urban practices at all five time points. The variance in requesting rates in rural GP practices was higher by 9-52% for electrolyte profiles, 280-460% for lipid profiles, 280-1212% for PSA, 224-367% for thyroid profiles, 29-82% for liver profiles, 2-301% for HbA_{1c} (except for the period 2012-13 when the variance was 16% lower in rural areas), 34-50% for ACR (except for the periods of 2011-12 and 2012-13 when the variance in rural areas was lower by 15% and 16% respectively), and 36-161% for immunoglobulins (except for the period of 2015-16 when the variance was 9% lower in rural areas). Furthermore, we found the rural-urban differences in variance for ordering rates of PSA and thyroid profiles statistically significant ($p_{FK} < 0.05$) (table 2)

Socioeconomic factors

We observed a very weak association between the Multiple Deprivation Measure (MDM) and requesting activity. Accordingly, the lowest Kendall τ of 0.01 was reported for HbA_{1c} while the highest τ = 0.18 for PSA (supplementary table S3). Similarly, the relationship between the Health Deprivation & Disability rank and the standardized number of test request was the lowest for HbA_{1c} (τ = 0.06) and highest for PSA (τ = 0.31). Again, the τ value indicated a weak correlation in both cases.

Table 2. The significance of differences in the distribution and variability in test request rates between GP practices located in rural and urban areas. p_{MWU}: Mann Whitney U *p*-value assessing differences in distributions of laboratory test rates between rural and urban practices. p_{FK}: Fligner-Killeen *p*-value referring to the significance level of differences in variances. For both tests, a $p < 0.05$ was considered significant (*).

	Electrolyte profiles	Lipid profiles	PSA	Thyroid profiles	HbA1c	Liver profiles	ACR	Immunoglobulins
p_{MWU}								
2011-2012	0.1	0.0004*	0.01*	0.009*	0.3	0.03*	0.4	0.2
2012-2013	0.08	0.001*	0.03*	0.01*	0.7	0.02*	0.5	0.4
2013-2014	0.8	0.01*	0.02*	0.07	1.0	0.3	0.5	0.4
2014-2015	0.6	0.01*	0.02*	0.1	0.9	0.3	0.5	0.03*
2015-2016	0.7	0.03*	0.003*	0.07	0.7	0.3	0.08	0.01*
p_{FK}								
2011-2012	0.4	0.5	0.03*	0.2	0.6	1	0.2	0.8
2012-2013	0.3	0.6	0.01*	0.3	0.9	0.4	0.7	0.3
2013-2014	0.2	0.6	0.06	0.03*	0.5	0.7	1	0.5
2014-2015	0.9	0.6	0.1	0.02*	0.5	0.9	0.3	0.6
2015-2016	0.7	0.7	0.2	0.2	0.2	0.7	0.8	0.1

Quality and Outcomes Framework indicators

To evaluate the relationship between standardized number of laboratory tests and QOF indicator scores, we looked at the management of diabetes. There are two main reasons for that. First, the guidelines for management of diabetes are widely used by practitioners to guide the care of their patients. Secondly, HbA_{1c} is a specific test in diabetes care and does not play an important role in the monitoring or diagnosis of any other condition.

The three target levels for HbA_{1c} (59, 64, and 75 mmol/mol) in the QOF were introduced to improve glycaemic control across the distribution of values. We investigated the overall QOF points under these 3 categories i.e. DM007 ('The percentage of patients with diabetes in whom the last HbA_{1c} is 59 mmol/mol or less in the preceding 15 months'), DM008 ('The percentage of patients with diabetes in whom the last HbA_{1c} is 64 mmol/mol or less in the preceding 15 months'), and DM009 ('The percentage of patients with diabetes in whom the last HbA_{1c} is 75 mmol/mol or less in the preceding 15 months'). We also combined the number of points for each of these 3 categories, calculated the overall achievement rate, and investigated its relationship with inter-practitioner variability in test requesting of HbA_{1c}.

All practices achieved the maximum 17 points available under QOF clinical indicator DM007. All but nine practices attained the maximum 8 points available under DM008 (range: 4.1-8.0) and 31 practices attained the maximum 10 points available under DM009 (range: 6.36-10.0). The strength of relationship between the number of HbA_{1c} tests performed and the GP practice effectiveness, as measured by the QOF overall achievement rate (combined DM007, DM008, and DM009) was very weak (τ = 0.12).

Discussion

We evaluated temporal changes in variability and number of laboratory tests ordered by individual GP practices in the WHSCT. We also investigated a range of key demographic, socioeconomic, geographical, and clinical factors to assess whether any of these factors are likely to explain a part of the observed variation in requesting activity.

While we found an overall decrease in the median request rates for lipid profiles, thyroid profiles, liver profiles, and ACR (statistically significant only for lipid profiles), we observed an increase in the median standardized number of PSA, HbA_{1c}, immunoglobulins, and electrolyte profiles (statistically significant only for PSA and HbA_{1c}). Furthermore, we identified substantial differences in variation of adjusted test request rates for individual tests, with an index of variability oscillating between 2.0-2.9 for electrolyte profiles, liver profiles, and HbA_{1c}, 2.8-3.3 for thyroid profiles, 3.1-3.7 for lipid profiles, 6.3-7.6 for PSA, and reaching up to 69.8 for immunoglobulins. Our finding of high levels of variability between practices is consistent with previous research.[6,10,28,29] This large variation may suggest that a portion of ordered tests had been not clinically relevant according to standards and guidelines and therefore, had limited or no benefit to the patient care.

Temporal changes in between practice variability of request rates were associated with a reduction in variance for electrolyte profiles (16.8%), lipid profiles (15.3%), thyroid profiles (28.3%), and liver profiles (23.5%) and the increase in variation of ordering rates for ACR (40%), immunoglobulins (272%), and HbA_{1c} (500%) over the period of investigation. The overall downward trend in variability of profile test request rates might have been related to the implementation of the demand optimization intervention in the WHSCT over the three year period from Apr 2012 to Mar 2015. The intervention

was developed to support appropriate use of laboratory testing in the WHSCT and consisted of several elements including educational materials on the benefits of the optimal tests utilization, information on minimal retest intervals, the review of the test requesting processes, and financial incentive. The 500% increase in variability of HbA_{1c} ordering rates, on the other hand, may have been caused by between practice differences in the adoption of HbA_{1c} as a diagnostic test (from 2012) and inconsistent implementation of new guidelines on appropriate rate of diabetes monitoring.

The inter-practitioner variations in test ordering were found unrelated to demographic and socioeconomic factors. We showed that distributions of patients' age and sex had little impact on ordering rates for a predetermined set of pathology tests. The socioeconomic status of GP practice also did not appear to identify low or high requestors of laboratory tests. No clear relationship between test ordering and age, gender or deprivation measures was reported by other studies.[6,30-32]

The practice location was found to be a significant determinant of variability in test use. Both the variability and median number of request rates were generally higher in GP practices located in rural areas at all time points. This probably reflects the differences in clinical practice associated with the specific aspects of practice organisation and workflow in rural and urban practices as well as personal characteristics of general practitioners. The practice location was previously related to the between-practice variation in prescribing.[33-35] Further studies are needed to explore in more detail the potential reasons for the rural-urban discrepancy in test utilization.

Finally, we found no evidence of a significant correlation between test requesting rates and clinical outcomes. This observation is consistent with other studies.[10] The fact that the majority of practices attained maximum points for the three target levels for

HbA_{1c} (59, 64, and 75 mmol/mol) implies that the glycaemic control was generally good and could not explain the variability in HbA_{1c} testing.

Implications for practice and research

The findings of our study suggest that there may be considerable potential for the rationalization of test ordering through minimizing the between practice variability in test utilization. Yet, it is important to establish whether the observed variation is, and to what extent, associated with over-requesting (unnecessary repeat requesting of tests) in GP practices with high ordering rates or on the contrary, it reflects a failure to prescribe clinically indicated tests by GP practices characterized by low use of tests. Further analysis on the degree and detection of inappropriate use of laboratory resources in primary care could contribute to improving the consistency, efficiency, and cost-effectiveness of patient diagnostics, monitoring, and treatment and hence, reduce unnecessary costs to patient care.

Further exploration of variations in requesting activity in the context of factors not considered in this study (e.g. practitioner-specific characteristics) may help identify and implement appropriate optimization strategies to manage demand for laboratory tests. Previous studies reported on the number of successful approaches in fostering best practice and reducing variation in test utilization including implementing locally agreed clinical guidelines, changing test order forms, incorporating clinical decision support tools with embedded retest interval rules, conducting audits of GPs on their request rates, and providing financial incentives.[36-42] In fact, several studies showed that multifaceted interventions were most successful in optimizing laboratory demand.[43,44]

Finally, it is essential to better understand implications of variability in laboratory test requesting for the cost and quality of care. In particular, an important question to be

answered is whether or not the growing costs associated with increased use of laboratory services has led to commensurate benefits to the patient.

Strengths and limitations

Test requesting data were directly extracted from the HSC Business Object system that captures information on tests' use from three clinical chemistry departments of Altnagelvin Area Hospital, Tyrone County Hospital, and South West Acute Hospital. Accordingly, our analysis was based on all available data regarding requesting activity in the N. Ireland Western Health and Social Care Trust (WHSCT) in the period of 1 Apr 2011 – 31 Mar 2016 and hence, was not subject to selection bias. Furthermore, our study provides a first comprehensive insight into the use of laboratory tests and factors accounting for the variation in between practice test utilization in the WHSCT primary care system.

Due to data unavailability, we were unable to investigate the relationship between laboratory use patterns and practitioner-specific characteristics including GP's age, education, and medical training. Such analysis could help identify potential reasons behind variation in clinical practice. Besides the lack of information on the practitioners, the present study was limited by a paucity of research evidence in this area. We were also unable to retrieve consolidated data on both gender and age of patients registered in individual GP practices and therefore, assess the combined effect of sex and age distributions on test requesting activity.

Conclusion

This study investigated the patterns and temporal changes in request rates across a range of frequently ordered laboratory tests. In addition, it explored potential determinants of the substantial inter-practice variability in the use of laboratory tests

and found that differences in requesting activity appear unrelated to either demographic and socioeconomic characteristics of GP practices or clinical outcome indicators. Our results highlight the need for further investigations to identify other potential factors that may account for the differences in test utilization between practitioners.

Acknowledgments

The authors would wish to thank Dr KongFatt Wong-Lin for helpful discussions and Graham Moore for computing and technical support.

References

1. Newman-Toker DE, McDonald KM, Meltzer DO, et al. How much diagnostic safety can we afford, and how should we decide? A health economics perspective. *BMJ Qual Saf.* 2013; 22: ii11-ii20.
2. Karakusevic S, Edwards N, Lewis R, et al. *The future of pathology services.* Nuffield Trust. 2016.
3. Smellie WSA. Demand management and test request rationalization. *Ann Clin Biochem* 2012; 49(4): 323-36.
4. Bareford D, Hayling A. Inappropriate use of laboratory services: long term combined approach to modify request patterns. *BMJ* 1990; 301(6764):1305-1307.
5. An Independent Review for the Department of Health. Report of the Review of NHS Pathology Services in England. Chaired by Lord Carter of Coles. 2006.
6. Smellie WSA, Galloway MJ, Chinn D, et al. Is clinical practice variability the major reason for differences in pathology requesting patterns in general practice? *Journal of Clinical Pathology*. 2002;55(4):312-314.
7. Sá L, Teixeira AS, Tavares F et al. Diagnostic and laboratory test ordering in Northern Portuguese Primary Health Care: a cross-sectional study. *BMJ Open*. 2017;7(11):e018509.
8. Kwok J, Jones B. Unnecessary repeat requesting of tests: an audit in a government hospital immunology laboratory. *Journal of Clinical Pathology*. 2005;58(5):457-462.
9. Vaidya B, Ukoumunne OC, Shuttleworth J, et al. Variability in thyroid function test requests across general practices in south-west England. Qual Prim Care. 2013;21: 143-8.
10. O'Kane MJ, Casey L, Lynch PM, et al. Clinical outcome indicators, disease prevalence and test request variability in primary care. *Ann Clin Biochem* 2011; 48(2): 155-158.

11. Smellie W, Galloway M, Chinn D. Benchmarking general practice use of pathology services: a model for monitoring change. *Journal of Clinical Pathology*. 2000;53(6):476-480.
12. Weydert JA, Nobbs ND, Feld R, et al. A simple, focused, computerized query to detect overutilization of laboratory tests. *Arch Pathol Lab Med* 2005;129(9):1141-1143.
13. Wong ET, McCarron MM, Shaw ST. Ordering of laboratory tests in a teaching hospital: can it be improved? *JAMA* 1983;249(22):3076-3080.
14. Young DW. Improving laboratory usage: a review. *Postgrad Med J* 1988;64(750):283-289.
15. Epstein AM, McNeil BJ. Relationship of beliefs and behavior in test ordering. *Am J Med* 1986; 80:865–870.
16. van Diepen S, Bakal JA, Lin M et al. Variation in critical care unit admission rates and outcomes for patients with acute coronary syndromes or heart failure among high- and low-volume cardiac hospitals. *J Am Heart Assoc* 2015; 4:e001708.
17. Adam MA, Goffredo P, Youngwirth L et al. Same thyroid cancer, different national practice guidelines: When discordant American Thyroid Association and National Comprehensive Cancer Network surgery recommendations are associated with compromised patient outcome. *Surgery* 2016; 159:41–50.
18. Northern Ireland Statistics and Research Agency. *Review of the Statistical Classification and Delineation of Settlements*. 2015. Available at: http://www.nisra.gov.uk/archive/geography/review-of-the-statistical-classification-and-delineation-of-settlements-march-2015.pdf. Accessed: Aug 2018.
19. Northern Ireland Statistics and Research Agency. *Northern Ireland Multiple Deprivation Measures 2017*. 2017. Available at: https://www.nisra.gov.uk/publications/nimdm17-results. Accessed: Jul 2018.
20. Department of Health. Quality and outcomes framework (QOF) achievement data 2015/16. 2016. Available at: https://www.health-ni.gov.uk/publications/quality-and-outcomes-framework-qof-achievement-data-201516. Accessed: Aug 2018

21. Department for Environment, Food & Rural Affairs. *Rural Urban Classification.* 2016. Available at: https://www.gov.uk/government/collections/rural-urban-classification. Accessed: Aug 2018.
22. Sen A, Srivastava M. *Regression analysis: theory, methods, and applications.* Springer Science & Business Media; 2012.
23. Jekel JF, Katz DL, Elmore JG et al. *Epidemiology, biostatistics and preventive medicine.* Elsevier Health Sciences; 2007.
24. Conover WJ, Johnson ME, Johnson MM. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*. 1981;23(4):351-61.
25. Li WK, Stanford DA, Yu H. *Advances in Time Series Methods and Applications.* Springer; 2016.
26. Gibbons JD, Fielden JD. *Nonparametric measures of association*. Sage. 1993
27. Gupta S, Verma M, Gupta AK, et al. Are we using Thyroid Function Tests Appropriately? *Indian Journal of Clinical Biochemistry* 2011;26(2):178-181.
28. Mogyorosy, Z, Mogyorosy, G. Practice pattern and geographic variation in test ordering: a literature review. *Orv Hetil* 2006;147:25–31.
29. Larsson, A, Palmer, M, Hulten, G, et al. Large differences in laboratory utilisation between hospitals in Sweden. *Clin Chem Lab Med* 2000;38:383–9.
30. Selvin, E, Steffes, MW, Zhu, H, Glycated haemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. *N Engl J Med* 2010;362:800–11.
31. Miyakis, S, Karamanof, G, Liontis, M, Mountokalakis, TD. Factors contributing to inappropriate ordering of tests in an academic medical department and the effect of an educational feedback strategy. Postgrad Med J 2006;82:823–9
32. Beckett, GJ, Toft, AD. First-line thyroid function tests – TSH alone is not enough. *Clin Endocrinol* 2003;58:20–1.
33. Bucholc M, O'Kane M, Ashe S, et al. Prescriptive variability of drugs by general practitioners. *PloS One* 2018;13(2):e0189599.

34. Morrison J, Anderson MJ, Sutton M, Munoz-Arroyo R, McDonald S, Maxwell M, et al. Factors influencing variation in prescribing of antidepressants by general practices in Scotland. *Br J Gen Pract* 2009;59(559): e25-e31.
35. Holm M, Olesen F. Factors affecting prescription of psychotropic drugs in general practice. *Scandinavian journal of primary health care* 1988;6(3):169-73.
36. Bugter-Maessen AM, Winkens RA, Grol RP, et al. Factors predicting differences among general practitioners in test ordering behaviour and in the response to feedback on test requests. *Fam Pract.* 1996;13: 254–258.
37. Baker R, Falconer Smith J, Lambert PC. Randomised controlled trial of the effectiveness of feedback in improving test ordering in general practice. *Scand J Prim Health Care.* 2003;21(4):219-223.
38. Buntinx F, Knottnerus JA, Crebolder HF, et al. Reactions of doctors to various forms of feedback designed to improve the sampling quality of cervical smears. *Qual Assur Health Care* 1992;4(2):161-166.
39. Bunting PS, Van Walraven C. Effect of a controlled feedback intervention on laboratory test ordering by community physicians. *Clin Chem.* 2004; 50(2):321-326.
40. Showstack JA, Schroeder SA, Matsumoto MF. Changes in the use of medical technologies, 1972–1977: a study of 10 inpatient diagnoses. *N Engl J Med.*1982;306: 706–712.
41. Pysher TJ , Bach PR , Lowichik A , et al . Chemistry test ordering patterns after elimination of predefined multitest chemistry panels in a children's hospital. *Pediatr Dev Pathol* 1999; 2:446–53.
42. Martins CM , da Costa Teixeira AS , de Azevedo LF , et al . The effect of a test ordering software intervention on the prescription of unnecessary laboratory tests - a randomized controlled trial. *BMC Med Inform Decis Mak* 2017; 17:20.
43. Oxman AD, Thomson MA, Davis DA, Haynes RB. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *Can Med Assoc J* 1995; 153:1423-31.

44. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993; 342:1317-22.

Contributors: MB and MJO had the original idea for this study. SA led the data collection. MB designed the methodology, performed the analysis, and drafted the manuscript. MJO, CM, and LM contributed to the drafting and critical revision of the manuscript.

Funding: This project was supported by the European Union's INTERREG VA Programme, managed by the Special EU Programmes Body (SEUPB).

Competing interests: None declared.

Ethics approval: Not applicable.

Data sharing statement: No additional data are available.

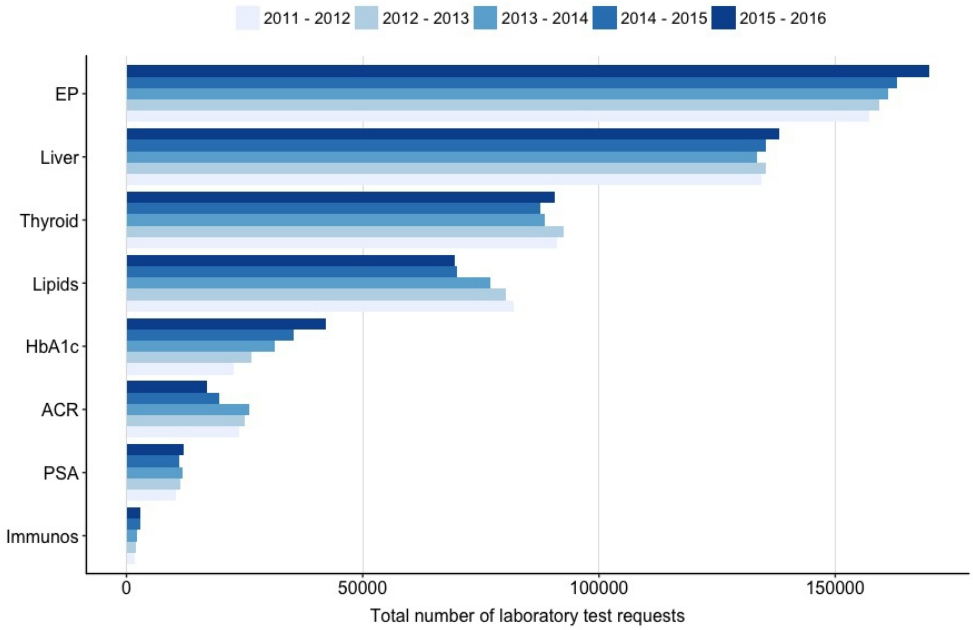
Figures

Figure 1. The total number of 8 considered clinical biochemistry tests/test groups ordered by 55 general practices during the period 1 Apr 2011 – 31 Mar 2016.

Figure 2. Temporal variability of the standardized laboratory test requests for (a) electrolyte profiles, (b) lipid profile, (c) prostate-specific antigen (PSA), (d) thyroid profiles, (e) liver profiles, (f) HbA_{1c}, (g) urine albumin/creatinine ratio (ACR), and (h) immunoglobulins (Immunos) for 55 considered general practices. Each data point (dot): a single practice. Solid, horizontal line inside the box: median. Lower and upper "hinges" of the boxplots: 1st and 3rd quartiles, respectively. Lower and upper extremes of whiskers: interval boundaries of the non-outliers (black dots). Data outside interval: outliers.

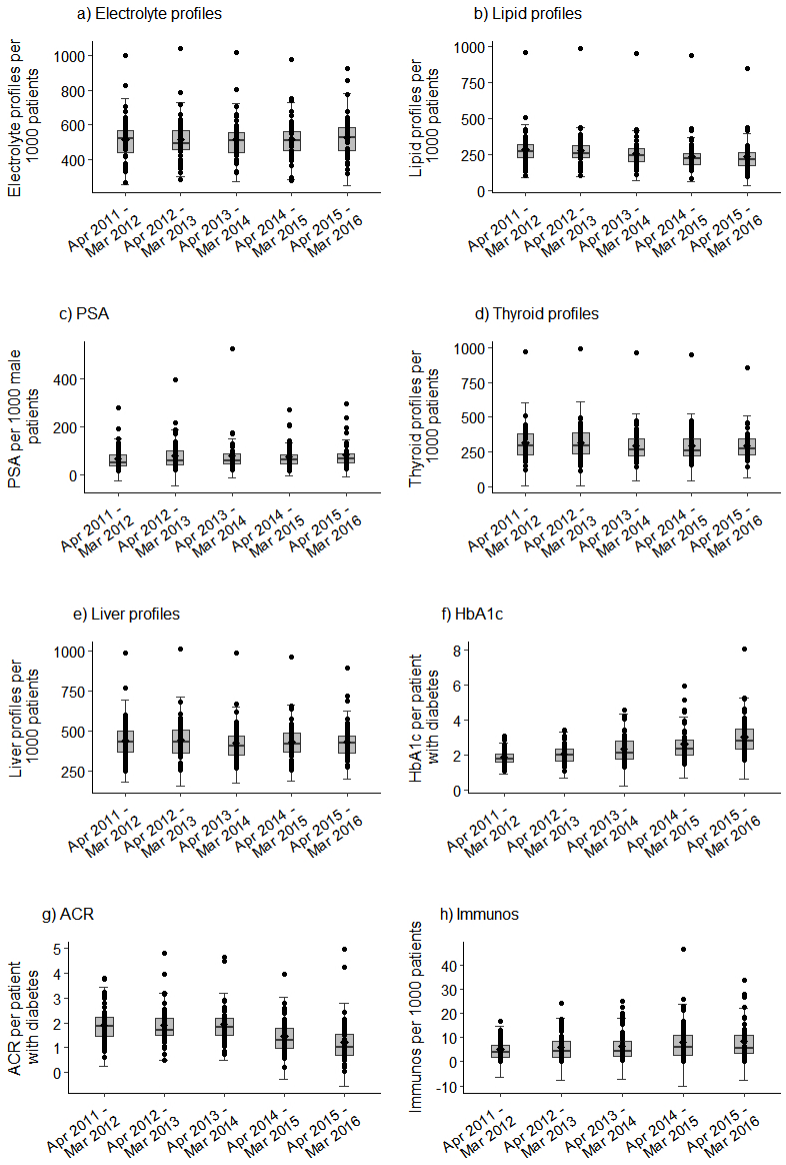
Figure 3. Trend lines for median (blue) and variance (red) of the standardized number of (a) electrolyte profiles, (b) lipid profile, (c) prostate-specific antigen (PSA), (d) thyroid profiles, (e) liver profiles, (f) HbA_{1c}, (g) urine albumin/creatinine ratio (ACR), and (h) immunoglobulins (Immunos).

Figure 4. The temporal median-variance relationship of the standardized number of laboratory test requests across the years. Circle/triangle: rural/urban general practices.

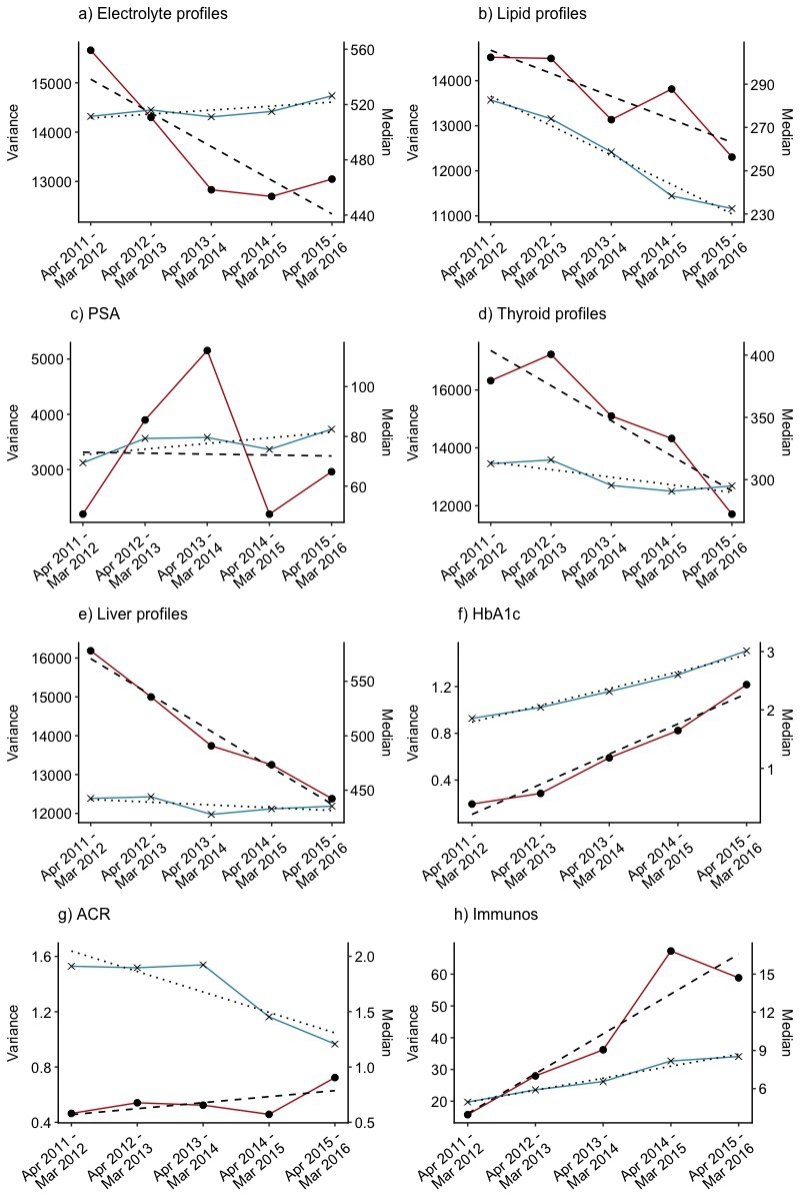


The total number of 8 considered clinical biochemistry tests/test groups ordered by 55 general practices during the period 1 Apr 2011 – 31 Mar 2016.

317x211mm (72 x 72 DPI)

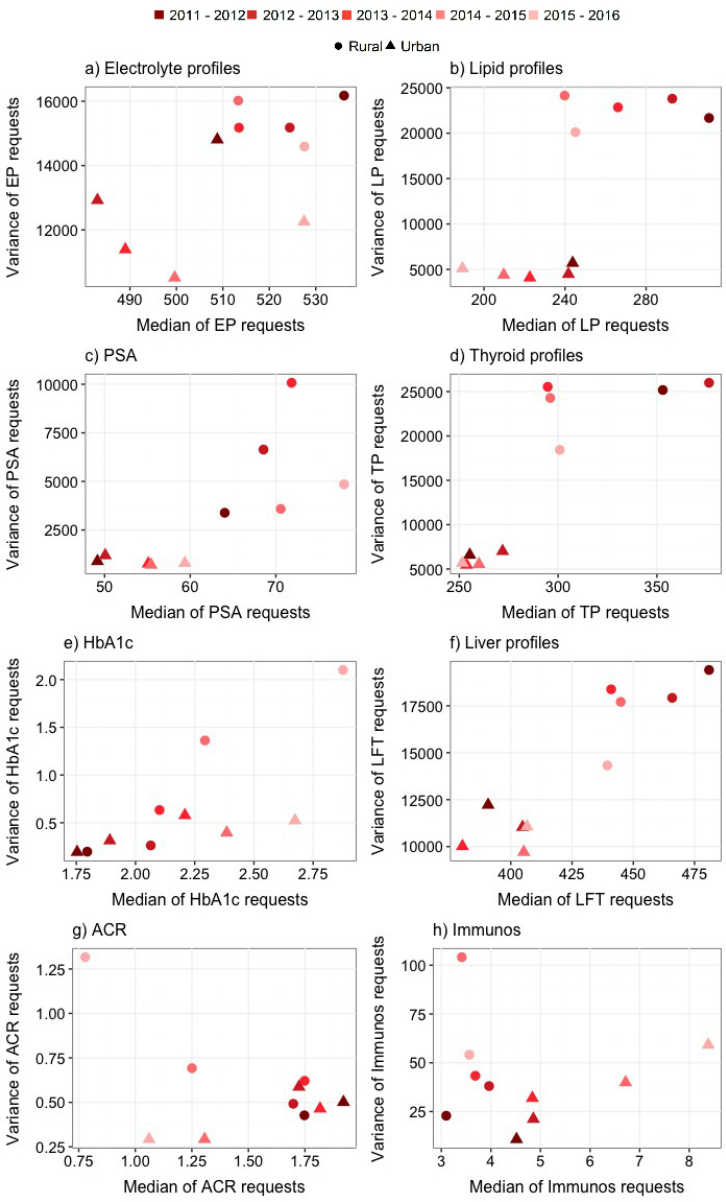


Temporal variability of the standardized laboratory test requests for (a) electrolyte profiles, (b) lipid profile, (c) prostate-specific antigen (PSA), (d) thyroid profiles, (e) liver profiles, (f) HbA1c, (g) urine albumin/creatinine ratio (ACR), and (h) immunoglobulins (Immunos) for 55 considered general practices. Each data point (dot): a single practice. Solid, horizontal line inside the box: median. Lower and upper "hinges" of the boxplots: 1st and 3rd quartiles, respectively. Lower and upper extremes of whiskers: interval boundaries of the non-outliers (black dots). Data outside interval: outliers.



Trend lines for median (blue) and variance (red) of the standardized number of (a) electrolyte profiles, (b) lipid profile, (c) prostate-specific antigen (PSA), (d) thyroid profiles, (e) liver profiles, (f) HbA1c, (g) urine albumin/creatinine ratio (ACR), and (h) immunoglobulins (Immunos).

282x423mm (72 x 72 DPI)



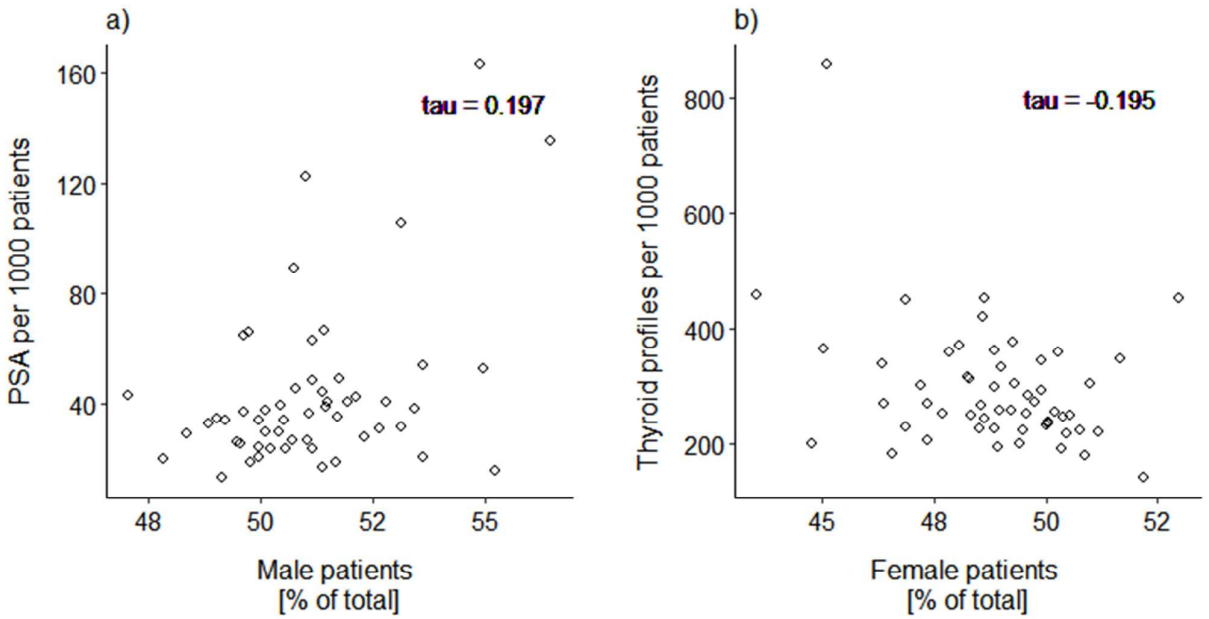
The temporal median-variance relationship of the standardized number of laboratory test requests across the years. Circle/triangle: rural/urban general practices.

246x405mm (72 x 72 DPI)

Supplementary Table S1.

	Electrolyte profiles	Lipid profiles	Thyroid profiles	Liver profiles
% >= 65 years old	0.2107	0.2321	0.08	0.1964

Supplementary Figure S2.



Supplementary Table S3.

	Electrolyte profiles	Lipid profiles	PSA	Thyroid profiles	HbA1c	Liver profiles	ACR	Immunoglobulins
Multiple Deprivation Measure	0.034	0.116	0.180	0.119	0.012	0.083	-0.094	-0.167
Health Deprivation & Disability	0.124	0.256	0.309	0.239	0.062	0.185	-0.021	-0.156

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	5-6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6-7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-8
Bias	9	Describe any efforts to address potential sources of bias	6-7,20
Study size	10	Explain how the study size was arrived at	6-7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6-8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8-9
		(*b*) Describe any methods used to examine subgroups and interactions	9
		(*c*) Explain how missing data were addressed	Not applicable
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	Not applicable
		(*e*) Describe any sensitivity analyses	Not applicable
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Not applicable
		(b) Give reasons for non-participation at each stage	Not applicable
		(c) Consider use of a flow diagram	Not applicable
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Not applicable
		(b) Indicate number of participants with missing data for each variable of interest	Not applicable
Outcome data	15*	Report numbers of outcome events or summary measures	Not applicable
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Not applicable
		(*b*) Report category boundaries when continuous variables were categorized	16
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	10-16
Discussion			
Key results	18	Summarise key results with reference to study objectives17	17-18
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	20
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	19
Generalisability	21	Discuss the generalisability (external validity) of the study results	17-18
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	26

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Primary care use of laboratory tests in Northern Ireland's Western Health and Social Care Trust: a cross-sectional study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-026647.R1
Article Type:	Research
Date Submitted by the Author:	27-Mar-2019
Complete List of Authors:	Bucholc, Magda; University of Ulster - Magee Campus, School of Computing, Engineering and Intelligent Systems OKane, Maurice; Altnagelvin Hospital, Clinical Chemistry Mullan, Ciaran; Health and Social Care Board Western Local Commissioning Group Ashe, Siobhan; Altnagelvin Area Hospital Maguire, Liam; University of Ulster - Magee Campus, School of Computing, Engineering and Intelligent Systems.
Primary Subject Heading:	Pathology
Secondary Subject Heading:	Evidence based practice, Research methods, Medical management, Health services research, General practice / Family practice
Keywords:	Pathology < BASIC SCIENCES, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PATHOLOGY, Chemical pathology < PATHOLOGY, PRIMARY CARE

Primary care use of laboratory tests in Northern Ireland's Western Health and Social Care Trust: a cross-sectional study.

Magda Bucholc[1], Maurice J O'Kane[2], Ciaran Mullan[3], Siobhan Ashe[2], Liam Maguire[1]

[1] Intelligent Systems Research Centre, University of Ulster, Magee Campus, Londonderry BT48 7JL, Northern Ireland, UK

[2] Altnagelvin Area Hospital, Western Health and Social Care Trust, Glenshane Road, Londonderry BT47 6SB, Northern Ireland, UK

[3] Western Local Commissioning Group, Gransha Park House, Clooney Road, Londonderry BT 47 6FN, Northern Ireland, UK

Correspondence to

Dr Magda Bucholc; m.bucholc@ulster.ac.uk

Abstract

Objectives To describe the laboratory test ordering patterns by general practitioners (GPs) in Northern Ireland Western Health and Social Care Trust (WHSCT) and explore demographic and socio-economic associations with test requesting.

Design Cross-sectional study.

Setting Western Health and Social Care Trust, Northern Ireland.

Participants 55 WHSCT primary care medical practices that remained open throughout the study period 1 April 2011 - 31 March 2016.

Outcomes To identify the temporal patterns of laboratory test ordering behaviour for 8 commonly requested clinical biochemistry tests/test groups in WHSCT. To analyse the extent of variations in laboratory test requests by GPs and to explore whether these variations can be accounted for by clinical outcomes or geographical, demographic, and socioeconomic characteristics.

Results The median number of adjusted test request rates over five consecutive years of the study period decreased by 45.7% for urine albumin/creatinine ratio (ACR) ($p < 0.000001$) and 19.4% for lipid profiles ($p < 0.000001$) while a 60.6%, 36.6%, and 29.5% increase was observed for HbA_{1c} ($p < 0.000001$), immunoglobulins ($p = 0.000007$), and PSA ($p = 0.0003$) respectively. The between-practice variation in test ordering rates increased by 272% for immunoglobulins ($p = 0.008$) and 500% for HbA_{1c} ($p = 0.0001$). No statistically significant relationship between ordering activity and either demographic (age and gender) and socioeconomic factors (deprivation) or Quality and Outcome Framework (QOF) scores was observed. We found the rural-urban differences in between-practice variability in ordering rates for lipid profiles, thyroid profiles, PSA, and immunoglobulins to be statistically significant at the Bonferroni-adjusted significance level $p < 0.01$.

Conclusions We explored potential factors of the inter-practice variability in the use of laboratory tests and found that differences in requesting activity appear unrelated to either demographic and socioeconomic characteristics of GP practices or clinical outcome indicators.

Strengths and limitations of this study

- The study provides a comprehensive analysis of temporal changes in laboratory test utilization patterns and establishes the extent of variability in test requesting activity across general practices in Northern Ireland's Western Health and Social Care Trust.
- The variation in test ordering, not related to demographic and socioeconomic characteristics of practices, practice location or clinical outcome indicators, may reflect inappropriate laboratory test utilization and hence, suggest a potential for more efficient demand management of laboratory services.
- Given a cohort of general practices within one catchment area, our results provide evidence of differences in behaviour of individual GPs when managing patients with similar clinical symptoms.
- Failure to collect and cross-tabulate data on characteristics of general practitioners (GPs), such as GP's age, years of experience, medical training was a study limitation and a missed opportunity in assessing the influence of practitioner factors on the variation in test ordering behaviour.

Introduction

Despite the important role of laboratory testing in the diagnosis and monitoring of disease, there is concern about the increasing number of requested tests and in particular, large differences in laboratory utilization between clinical teams.[1] In the UK, laboratory test orders grew by approximately 5% per year in recent years and inappropriate test requests are considered to be an important cause for this increase[2,3]. Although pathology expenditures account for only 5–6% of the UK total health budget, they are viewed as a potential source of savings, most likely because the costs can be easily identified and measured.[4] According to the Department of Health, the rationalization of pathology services, including demand management of laboratory tests and elimination of unnecessary requesting, could produce savings of at least £500 million.[5]

Unnecessary testing is not only wasteful of resources but impacts on patients directly through the requirement for venepuncture and the follow up of minor (and possibly insignificant) abnormalities which may cause patient anxiety. On the other hand, inappropriate under requesting may cause harm through failure to diagnose or manage disease optimally. Several studies suggested that unnecessary and inappropriate utilization of laboratory services is closely linked to inter-practitioner variability in test requesting.[6-8] Despite the increased availability of clinical management guidelines promoting harmonization of the use of laboratory tests, there is still substantial variation in test utilization among general practitioners.[7,9] These differences appear to be unrelated to demographic characteristics of patient populations, socio-economic status of GP practices, disease prevalence or clinical outcome indicators.[6,7,10,11] Even if some of these variables have been shown to have an effect on test ordering patterns, the variation in requesting rates is so large that it can only be explained by differences in

attitudes towards the use of laboratory tests of individual practitioners.[10] Accordingly, factors such as confidence in clinical judgement, clinical experience, an attitude about clinical practice guidelines, a lack of knowledge regarding the correct use of tests and fear of litigation have been identified as potential sources of practice variation.[12-15]

Since unwarranted variation can lead to suboptimal clinical outcomes,[16,17] identification of factors contributing to differences in test requesting can provide useful information for optimising utilization of laboratory services. The aim of our study was to establish the extent of variability in test requesting and characterise temporal changes in test ordering patterns across general practices within the catchment area of the Northern Ireland (NI) Western Health and Social Care Trust (WHSCT) for a range of most commonly requested clinical biochemistry tests/test groups. In addition, we investigated potential factors associated with inter-doctor variability in the use of laboratory tests including geographical, demographic, and socioeconomic factors as well as Quality and Outcome Framework (QOF) scores.

Materials and Methods

Study design and data sources

We conducted a cross-sectional study of laboratory test ordering activity across general practices in the WHSCT in the period from 1 April 2011 to 31 March 2016. The data on the use of laboratory tests were obtained from a HSC Business Object XI clinical information system. We investigated requesting rates for 8 clinical biochemistry tests/test groups including electrolyte profile, thyroid profile (FT4 and TSH), liver profile, lipid profile, urine albumin/creatinine ratio (ACR), glycosylated haemoglobin (HbA_{1c}), prostate-specific antigen (PSA), and immunoglobulins. The profile tests i.e., electrolyte, lipid, liver, and thyroid profiles contained a number of different related analytes.

Individual elements of the profile tests could not be requested separately. Other tests, such as HbA_{1c} comprised one single analyte. The standardized number of laboratory tests requested in each general practice was determined by dividing the total number of requested tests by the number of registered patients and expressed as requests per 1000 patients, with the exception of HbA_{1c} and ACR for which ordering rates were expressed as tests per patient with diabetes and PSA rates, calculated per 1000 male patients. GP practice list size data (including the number of patients with diabetes) and patient demographic data was extracted from the Business Services Organisation (BSO) Family Practitioner Service Information and Registration Unit system. The rural-urban distribution of GP practices was based on the data from the Census Office of the Northern Ireland Statistics and Research Agency (NISRA).[18] Socioeconomic characteristics of GP practices were determined using the NISRA Neighbourhood Information System (NINIS).[19] QOF scores for individual practices were extracted from the website of the Northern Ireland Department of Health.[20]

Participants and setting

Data on laboratory tests requested from 55 general practices within the WHSCT were collected from the laboratory databases of Clinical Chemistry departments of the Altnagelvin Area Hospital, Tyrone County Hospital, and the Erne Hospital (subsequently the South West Acute Hospital). WHSCT provides health and social care services throughout the west of Northern Ireland, across the council areas of Derry City and Strabane District Council, Limavady in the Causeway Coast and Glens Borough Council, and Fermanagh and Omagh District Council.

Inclusion criteria

We examined laboratory test requests from 55 separate primary care medical practices within the catchment area of WHSCT that remained open throughout the study period i.e. during five consecutive years: Apr 2011 – Mar 2012, Apr 2012 – Mar 2013, Apr 2013 – Mar 2014, Apr 2014 – Mar 2015, and Apr 2015 – Mar 2016. To ensure the completeness and consistency of the data, test orders from WHSCT GP practices that were closed (2 GP practices) or taken over by other practices (1 GP practice) during the period of investigation were not taken into account.

Variables and characteristics

Our analysis was limited to 8 most frequently requested clinical biochemistry tests/test groups including electrolyte profile, thyroid profile, liver profile, lipid profile, ACR, HbA_{1c}, PSA, and immunoglobulins. All considered tests/profiles (with the exception of immunoglobulins and ACR) were listed on a paper pathology request form and ordered by ticking a box adjacent to the test name; requests for ACR and immunoglobulins were made by writing the test name in an open text field.

Rates of test requests were analysed in the context of geographical and socioeconomic characteristics of practices, patient demographics (age and gender), and clinical outcome indicators. Note that we were not able to obtain a consolidated data on both gender and age of patients (i.e., the number of patients split into age bands and broken down by gender) registered in individual GP practices.

The practice setting (rural v. urban area) was determined using the urban-rural classification of the Department for Environment, Food and Rural Affairs (DEFRA)[21] while the size of settlements was obtained from the NISRA Census Office.[18] A GP practice was defined as rural if its physical address was situated in a settlement of less

than 10,000. Accordingly, we classified 24 practices as rural and 31 as urban. Furthermore, GP practices were categorised based on the Northern Ireland Multiple Deprivation Measure (MDM) identified for individual Super Output Areas (SOAs).[19] The MDM comprises a weighted combination of 7 component measures (Income Deprivation; Employment Deprivation; Health Deprivation and Disability; Education, Skills and Training Deprivation; Proximity to Services; Living Environment; and Crime and Disorder) and ranges from 1 (most deprived SOAs) to 890 (least deprived SOAs). The Health Deprivation & Disability rank (one of the MDM domains)[19] was analysed separately as a potential factor contributing to inter-doctor variability in the use of laboratory tests.

Given a detailed breakdown of the age of patients on the individual practice lists, we examined the relationship between the percentage of patients over age 65 and ordering rates for electrolyte, thyroid, liver, and lipid profiles. In addition, we investigated the link between gender and requesting activity for PSA (for males) and thyroid profiles (for females).

To evaluate the impact of test requesting patterns for diabetes mellitus on the QOF clinical indicator scores in diabetes, we used the overall QOF points under clinical indicator DM007 ('The percentage of patients with diabetes in whom the last HbA_{1c} is 59 mmol/mol or less in the preceding 15 months'), DM008 ('The percentage of patients with diabetes in whom the last HbA_{1c} is 64 mmol/mol or less in the preceding 15 months'), and DM009 ('The percentage of patients with diabetes in whom the last HbA_{1c} is 75 mmol/mol or less in the preceding 15 months'). Specifically, we combined the number of points for each of these 3 categories, calculated the overall achievement rate, and investigated its relationship with inter-practitioner variability in test requesting of HbA_{1c}.

Outcome measures

Our outcome of interest was to identify the presence and extent of variations in primary care laboratory tests ordering and to evaluate temporal changes in both the standardised number of test requests and between-practice variability in requesting. In addition, we studied demographic, socio-economic, geographical, and clinical factors that may explain this variation.

Statistical analysis

Inter-practice variability in test requests was assessed by calculating the variance (σ^2). Furthermore, we computed 'variability index' (Var_i) defined as the top decile divided by the bottom decile of standardized test request rates.[10] Var_i is a dimensionless measure of dispersion allowing us to compare the amount of variation in request rates of individual tests despite their differences in scale. The Shapiro-Wilk test of normality was used to determine if the distribution of test ordering data deviated from a normal distribution.[22] Since the distribution of laboratory test request rates was found to be non-normal, the non-parametric statistics were implemented to perform further analysis. Mann Whitney U (MWU) test was employed to compare distributions of laboratory test rates between GP practices located in rural and urban areas.[23] The homogeneity of variances for requesting activity in rural and urban GP practices was assessed with the Fligner-Killeen (FK) test.[24] Significance of temporal changes in median and variation in test requesting rates was examined with the Mann–Kendall (MK) test.[25] To test the relationship between adjusted requesting rates and 1) Multiple Deprivation Measure; 2) Health Deprivation & Disability Deprivation rank; 3) proportions of patients over age 65; 4) distribution of patient's gender; and 5) QOF clinical indicator scores, we calculated the Kendall rank correlation coefficient (τ) with the corresponding p-value.[26] The Kendall coefficient (-1 ≤

$r \leq 1$) measures the strength of a correlation between two variables and assesses the degree of overall correspondence of variables' ranking.[26]

Since the test statistics were used to simultaneously evaluate the significance of observations for different case scenarios (i.e., differences over time or differences between clinical biochemistry tests/test groups), we took a conservative approach in selecting the significance level α by applying the Bonferroni correction.[23] This multiple comparison correction technique minimizes the risk of obtaining false positive results by using an adjusted α for each single test. Single threshold values α were calculated in such a way that the family wise error probability p_{fwe} (here adopted $p_{fwe} < 0.05$) was retained at the global level.[23] Accordingly, the significance level α for a single test was approximated by dividing the global error probability p_{fwe} by the total number of independent tests.[23]

Patient and public involvement

No patients or general practitioners were involved in defining the research question or outcome measures nor were they engaged in the design or implementation this study. The study was approved by the Western Health and Social Care Trust (WHSCT). All identifiable information used for the purpose of the study was anonymised and not traceable to individual patients or general practitioners.

Results

Test ordering patterns

We analysed the laboratory test request rates of 55 general practices within the catchment area of the WHSCT comprising a total of 316 382 (2011-12), 316 688 (2012-13), 318 057 (2013-14), 319 383 (2014-2015), and 326 429 (2015-16) patients. Figure

1 shows the total number of 8 considered clinical biochemistry tests/test groups ordered during the study period of five consecutive years. The total number of ordered tests was 523 111 (2011-12), 531 849 (2012-13), 531 583 (2013-14), 525 146 (2014-2015), and 542 118 (2015-16) with electrolyte profiles found to be the most frequently ordered tests, making up approximately 30% of all requests in each year.

The median number of adjusted test request rates between the 2011-12 and 2015-16 decreased by 45.7% for urine albumin/creatinine ratio (ACR) ($p < 0.000001$) and 19.4% for lipid profiles ($p < 0.000001$). We observed a 60.6%, 36.6%, and 29.5% increase for HbA_{1c} ($p < 0.000001$), immunoglobulins ($p = 0.000007$), and PSA ($p = 0.0003$) respectively. The Bonferroni adjusted probability $p < 0.01$ was used in testing for statistical significance (table 1).

The between practice variation in the number of laboratory test requests is shown in Table 1. In the considered period, the Var_i was lowest for electrolyte profiles (2.0-2.3), liver profiles (2.2-2.5), and HbA_{1c} (2.2-2.9) and highest for immunoglobulins (19.2-69.8).. The number of outlier GP practices, i.e. practices with outlier ordering rates of laboratory tests contributed to the overall variability (figure 2). The number of requested tests for thyroid profiles, HbA_{1c} , and immunoglobulins per patient was over 3 times higher in the outlier GP practices (table 2). We however acknowledge that a statistical outlier in terms of test utilization is not necessarily an example of inappropriate practice. When compared to 2011-12, the variance in test utilization was 16.8% lower for electrolyte profiles, 15.3% for lipid profiles, 28.3% for thyroid profiles, and 23.5% for liver profiles in 2015-16. For ACR, immunoglobulins, and HbA_{1c}, the variance was 40%, 272%, and 500% higher respectively.

Table 1. Overview of general statistics

	EP	Lipids	PSA	Thyroid	HbA1c	Liver	ACR	Immunos
2011-2012								
Median, CI	520.5, (465.5, 545.0)	268.9, (242.2, 300.5)	53.2, (48.3, 79.8)	296.7, (266.2, 339.3)	1.8, (1.7, 2.0)	435.7, (394.3, 473.9)	1.9, (1.6, 2.1)	4.1, (3.1, 6.0)
Range	267 - 1002	101.8 - 954.5	19.6 - 279.3	121 - 969.4	1.1 - 3.1	254 - 989.1	0.6 - 3.8	0.3 - 16.8
Variance	15660.2	14519.1	2193.8	16320.0	0.2	16187.4	0.5	15.8
Var_i	2.3	3.4	7.0	3.3	2.2	2.5	3.4	18.3
2012-2013								
Median, CI	491.7, (474.6, 542.5)	257.4, (236.1, 288.5)	59.2, (51.5, 88.3)	291.7, (266.5, 342.1)	2.0, (1.8, 2.2)	433.7, (399.0, 473.4)	1.7, (1.6, 2.0)	4.5, (3.6, 7.9)
Range	283.6 - 1041.8	105.8 - 986.3	19.9 - 396.1	115.5 - 995.2	1.1 - 3.4	288.7 - 1016.4	0.5 - 4.8	0 - 24
Variance	14299.4	14494.8	3896.4	17230.2	0.3	14997.6	0.5	28.0
Var_i	2.2	3.4	7.6	3.3	2.4	2.4	3.7	69.8
2013-2014								
Median, CI	508.8, (469.2, 534.2)	239.3, (221.3, 271.2)	59.5, (54.9, 83.5))	264.3, (246.4, 318.0)	2.1, (1.9, 2.5)	413.8, (381.6, 453.3)	1.8, (1.6, 2.1)	4.4, (3.7, 8.4)
Range	325.1 - 1014.9	111 - 950.7	23.1 - 527.6	141.5 - 962.2	1.3 - 4.6	267.1 - 987.8	0.7 - 4.7	0.5 - 24.8
Variance	12831.5	13134.3	5153.6	15099.2	0.6	13739.4	0.5	36.2
Var_i	2.0	3.1	7.2	3.0	2.8	2.3	3.3	19.2
2014-2015								
Median, CI	508.7, (476.5, 541.4)	224.0, (198.6, 248.4)	63.9, (55.6, 81.0)	260.4, (244.9, 311.8)	2.3, (2.2, 2.8)	422.4, (389.5, 458.4)	1.3, (1.2, 1.7)	5.9, (4.7, 9.9)
Range	274.9 - 976.9	82.8 - 939.5	17.1 - 274	152.7 - 951.7	1.5 - 5.9	260.7 - 968	0.2 - 4	0 - 46.6
Variance	12696.9	13814.0	2192.6	14324.6	0.8	13251.9	0.5	67.3
Var_i	2.1	3.5	6.3	3.0	2.9	2.3	4.7	33.6
2015-2016								
Median, CI	527.5, (482.9, 556.9)	216.6, (193.2, 244.6)	68.9, (60.2, 86.7)	268.4, (251.6, 315.1)	2.8, (2.6, 3.3)	426.3, (391.7, 457.7)	1.0, (0.9, 1.3)	5.6, (5.0, 10.4)
Range	316 - 925.3	96.9 - 846.5	26.4 - 296.9	142.5 - 857.2	1.7 - 8.1	274.6 - 898.3	0.1 - 5	0 - 33.9
Variance	13048.3	12303.9	2961.0	11708.1	1.2	12378.9	0.7	58.9
Var_i	2.0	3.7	6.5	2.8	2.9	2.2	9.4	24.5

Table 2. Number of requested tests per patients in outlier (O) and non-outlier (NO) GP practices. The requesting rates for: 1) electrolyte profiles, liver profiles, thyroid profiles, lipid profiles, and immunoglobulins are expressed as the number of tests per patient; 2) PSA – as the number of tests per male patient; 3) ARC and HbA_{1c} – as the number of tests per patient with diabetes.

Period	Electrolyte profiles		Lipid profiles		PSA		Thyroid profiles		HbA_{1c}		Liver profiles		ACR		Immunos	
	O	NO	O	NO	O	NO	O	NO	O	NO	O	NO	O	NO	O	NO
2011-2012	0.88	0.49	0.65	0.25	0.22	0.06	0.97	0.28	3.04	1.78	0.84	0.42	1.80	1.83	0.017	0.005
2012-2013	0.58	0.50	0.99	0.25	0.28	0.07	1.00	0.29	3.32	1.99	1.02	0.42	1.61	1.80	0.024	0.006
2013-2014	0.85	0.50	0.95	0.24	0.22	0.07	0.96	0.28	4.59	2.32	0.74	0.41	1.50	1.84	0.024	0.006
2014-2015	0.78	0.50	0.52	0.21	0.22	0.07	0.95	0.27	5.00	2.46	0.97	0.42	1.96	1.43	0.034	0.008
2015-2016	0.87	0.51	0.51	0.21	0.23	0.07	0.86	0.27	8.06	2.93	0.74	0.41	1.55	1.13	0.027	0.008

Trends in the number and variation of test requests over time

We observed no statistically significant linear trend in the median standardized number of laboratory tests over the period of investigation as indicated by the Mann-Kendal test *p*-values (figure 3). Furthermore, no linear trend was found in the annual changes in variation in test requesting rates (figure 3). Note that to control the false discovery rate in multiple comparisons (for 8 laboratory tests), the Bonferroni adjusted probability $p < 0.00625$ was used in testing for statistical significance.

Test ordering variation between general practices and explanatory factors

The relationship between test requesting rates and potential explanatory factors was established based on the information on the number of ordered tests, patient demographics, and Quality and Outcomes Framework indicator scores obtained for the period 1 Apr 2015 – 31 Mar 2016.

Demographic characteristics of patient population

Proportion of the oldest age category of patients (> 75) constituted a relatively small group in each GP practice (mean = 6.1%, 95%CI = (5.7%,6.5%)). Hence, we combined the 65-74 and > 75 age categories to create a more meaningful group of patients in the older age band. We examined four tests for which differences in proportions of older patients were expected to be reflected in test requesting i.e. electrolyte profiles, liver profiles, lipid profiles, and thyroid profiles. Since we were unable to obtain a consolidated data on both gender and age of patients, we did not assess the relationship between the PSA requesting rates and the category of males aged 65 and over. We found no statistically significant relationship between the adjusted request rates of selected laboratory tests and the percentage distribution of patients of age over 65 (supplementary table S1).

Since previous studies reported on a higher rate of testing of thyroid hormones in females,[27] we looked at the strength of a relationship between the percentage of females in individual GP practices and requesting rates for thyroid profiles. We found no statistically significant association between these two characteristics (τ = -0.19, p = 0.05) (supplementary figure S1).

In addition, we examined the effect of gender distribution on the standardized number of PSA tests, with the PSA ordering rates expected to be higher for GP practices with higher percentage of males (supplementary figure S2). No statistically significant correlation between these two variables was found (τ = 0.15 and p = 0.1). Note that we acknowledge the fact that a combined effect of sex and age distributions might have had a more significant effect on PSA requesting activity. However, the restricted access to such data limited the opportunity of more insightful analysis.

Practice setting

Figure 4 shows a temporal median-variance relationship of the standardized number of laboratory test requests for rural and urban areas. The median number of tests ordered annually by practices located in rural areas was higher by approximately 27-37% for PSA, 14-30% for lipid profiles, 14-38% for thyroid profiles, and 8-23% for liver profiles. For ACR and immunoglobulins the median requesting rates were lower in rural areas by 1-27% and 18-57% respectively.

Across five consecutive time periods, the Mann Whitney U (MWU) test showed significant differences between 'rural' and 'urban' distributions of laboratory test rates for lipid profiles (p_{MWU} = 0.0004 in 2011-12, 0.001 in 2012-2013, and 0.0097 in 2013-14), thyroid profiles (p_{MWU} = 0.009 in 2011-12), PSA (p_{MWU} = 0.003 in 2015-2016), and immunoglobulins (p_{MWU} = 0.0096 in 2015-16) while no statistically significant rural-urban difference in between-practice variation in ordering rates was found for any of the studied

Table 3. The significance of differences in the distribution and variability in test request rates between GP practices located in rural and urban areas. p_{MWU}: the Mann Whitney U *p*-value assessing differences in distributions of laboratory test rates between rural and urban practices. p_{FK}: the Fligner-Killeen *p*-value referring to the significance level of differences in variance. In both cases, the Bonferroni-corrected $p < 0.01$ was considered statistically significant (*). Bonferroni-corrected significance cut-off of 0.01 was established by dividing the family wise error probability $p_{fwe} < 0.05$ by the number of tested associations i.e. 5 measurements per time.

	Electrolyte profiles	Lipid profiles	PSA	Thyroid profiles	HbA1c	Liver profiles	ACR	Immunoglobulins
p_{MWU}								
2011-2012	0.1	0.0004*	0.01	0.009*	0.3	0.03	0.4	0.2
2012-2013	0.08	0.001*	0.03	0.01	0.7	0.02	0.5	0.4
2013-2014	0.8	0.0097*	0.02	0.07	1.0	0.3	0.5	0.4
2014-2015	0.6	0.01	0.02	0.1	0.9	0.3	0.5	0.03
2015-2016	0.7	0.03	0.003*	0.07	0.7	0.3	0.08	0.0096*
p_{FK}								
2011-2012	0.4	0.5	0.03	0.2	0.6	1	0.2	0.8
2012-2013	0.3	0.6	0.01	0.3	0.9	0.4	0.7	0.3
2013-2014	0.2	0.6	0.06	0.03	0.5	0.7	1	0.5
2014-2015	0.9	0.6	0.1	0.02	0.5	0.9	0.3	0.6
2015-2016	0.7	0.7	0.2	0.2	0.2	0.7	0.8	0.1

tests (table 3). Note that p_{MWU} and p_{FK} lower than Bonferroni-corrected significance cut-off $\alpha < 0.01$ (corrected for 5 different measurements over time) were considered statistically significant.

Socioeconomic factors

We observed no statistically significant relationship between the Multiple Deprivation Measure (MDM) and test requesting activity (supplementary table S2). Similarly, the relationship between the Health Deprivation & Disability rank and the standardized number of test requests was found insignificant for any of the considered tests (supplementary table S2). Note that to adjust for multiple comparisons at a family-wise simultaneous error rate of $p_{fwe} < 0.05$, the Bonferroni-adjusted $\alpha < 0.003125$ (corrected for 16 different comparisons) was used to test for statistical significance.

Quality and Outcomes Framework indicators

To evaluate the relationship between standardized number of laboratory tests and QOF indicator scores, we looked at the management of diabetes. There are two main reasons for that. First, the guidelines for the management of diabetes are widely available to practitioners to improve clinical practice and care of diabetic patients. Secondly, HbA_{1c} is a specific test in diabetes care and does not play an important role in the monitoring or diagnosis of any other condition.

All practices achieved the maximum 17 points available under QOF clinical indicator DM007. All but nine practices attained the maximum 8 points available under DM008 (range: 4.1-8.0) and 31 practices attained the maximum 10 points available under DM009 (range: 6.36-10.0). The relationship between the number of HbA_{1c} tests performed and the GP practice effectiveness, as measured by the QOF overall

achievement rate (combined DM007, DM008, and DM009), was found statistically insignificant (τ = 0.12, p = 0.2).

Discussion

We evaluated temporal changes in variability and number of laboratory tests ordered by individual GP practices in the WHSCT. We also investigated a range of key demographic, socioeconomic, geographical, and clinical factors to assess whether any of these factors are likely to explain a part of the observed variation in requesting activity.

Inter-practice variability in test requests expressed by the index of variability was the lowest for profile tests and the highest for HbA_{1c} and immunoglobulins. This observation was consistent with previous studies.[10,28,29] In addition, we observed a 23.3% decrease in profile tests (electrolyte, lipid, liver, and thyroid profiles combined) between 2011-12 and 2015-16 (p = 0.009, the Bonferroni adjusted α < 0.01 used). This might have been related to the implementation of the demand optimization intervention in the WHSCT over the three year period from Apr 2012 to Mar 2015. The intervention was developed to support appropriate use of laboratory testing in the WHSCT and consisted of several elements including educational materials on the benefits of the optimal tests utilization, information on minimal retest intervals, the review of the test requesting processes, and financial incentive. Educational material covered the major clinical indications for a range of most commonly requested tests (including profile tests) and was circulated electronically to all GPs. The educational material was supplemented by face-to-face educational sessions with primary care teams and presentation of data showing the local variability on test requesting rates. Furthermore, all primary care teams were asked to engage in the process of reviewing test requesting procedures within their practice (i.e. what staff are allowed to request tests and what is the process for test requesting), and to reflect on the information provided on their practice test requesting rates and ranking

in comparison to other practices. GPs also received a very small incentive to participate in the scrutiny of their requesting processes and activities to reflect the extra time that this involved. The 500% increase in variance of HbA_{1c} ordering rates (statistically significant at the Bonferroni adjusted $\alpha < 0.01$), on the other hand, may have been caused by between practice differences in the adoption of HbA_{1c} as a diagnostic test (from 2012) and inconsistent implementation of new guidelines on appropriate rate of diabetes monitoring. This may indicate that recommended guidelines did not predispose GPs to change their perceptions on the value and role of HbA_{1c} test in patient assessment.

The inter-practitioner variations in test ordering were found unrelated to demographic and socioeconomic factors. We showed that distributions of patients' age and sex had little impact on ordering rates for a predetermined set of pathology tests. The socioeconomic status of GP practice also did not appear to identify low or high requestors of laboratory tests. No significant relationship between test ordering and age, gender or deprivation measures was reported by other studies.[6,30-32]

The practice location was found to be a significant factor associated with variability in use of lipid profiles, thyroid profiles, PSA, and immunoglobulins.. This may reflect the differences in clinical practice associated with the specific aspects of practice organisation and workflow in rural and urban practices as well as personal characteristics of general practitioners. The practice location was previously related to the between-practice variation in prescribing.[33-35] Further studies are needed to explore in more detail the potential reasons for the rural-urban discrepancy in test utilization, especially that these differences appear to be significant only for some clinical biochemistry tests/test groups

Finally, we found no evidence of a significant correlation between test requesting rates and clinical outcomes. This observation is consistent with other studies.[10] The fact that the majority of practices attained maximum points for the three target levels for HbA_{1c} (59, 64, and 75 mmol/mol) implies that the glycaemic control was generally good and could not explain the variability in HbA_{1c} testing.

Note that we took a conservative approach in determining the significance cut-off *α* when evaluating the significance of observations for multiple comparisons (e.g., differences in laboratory test requests over time or differences in request rates between considered clinical biochemistry tests/test groups) by applying the Bonferroni correction. We are aware that while the Bonferroni adjustment decreases the risk of Type I errors i.e., a probability of a false significant result, this happens at cost of inflating Type II errors i.e., the probability of accepting the null hypothesis when the alternative is true. However, by applying this correction technique, we believe we reinforced confidence in the significance of our results.

Implications for practice and research

The findings of our study suggest that there may be considerable potential for the rationalization of test ordering through minimizing the between practice variability in test utilization. Yet, it is important to establish whether the observed variation is, and to what extent, associated with over-requesting (unnecessary repeat requesting of tests) in GP practices with high ordering rates or on the contrary, it reflects a failure to prescribe clinically indicated tests by GP practices characterized by low use of tests. Further analysis on the degree and detection of inappropriate use of laboratory resources in primary care could contribute to improving the consistency, efficiency, and cost-effectiveness of patient diagnostics, monitoring, and treatment and hence, reduce unnecessary costs to patient care.

Further exploration of variations in requesting activity in the context of factors not considered in this study (e.g. practitioner-specific characteristics) may help identify and implement appropriate optimization strategies to manage demand for laboratory tests. Previous studies reported on the number of successful approaches in fostering best practice and reducing variation in test utilization including implementing locally agreed clinical guidelines, changing test order forms, incorporating clinical decision support tools with embedded retest interval rules, conducting audits of GPs on their request rates, and providing financial incentives.[36-42] In fact, several studies showed that multifaceted interventions were most successful in optimizing laboratory demand.[43,44]

Finally, it is essential to better understand implications of variability in laboratory test requesting for the cost and quality of care. In particular, an important question to be answered is whether or not the growing costs associated with increased use of laboratory services has led to commensurate benefits to the patient.

Strengths and limitations

Test requesting data were directly extracted from the HSC Business Object XI clinical information system that captures information on tests' use from three clinical chemistry departments of Altnagelvin Area Hospital, Tyrone County Hospital, and South West Acute Hospital. Accordingly, our analysis was based on all available data regarding requesting activity in the N. Ireland Western Health and Social Care Trust (WHSCT) in the period of 1 Apr 2011 – 31 Mar 2016 and hence, was not subject to selection bias. Furthermore, our study provides a first comprehensive insight into the use of laboratory tests and factors accounting for the variation in between practice test utilization in the WHSCT primary care system.

Due to data unavailability, we were unable to investigate the relationship between laboratory use patterns and practitioner-specific characteristics including GP's age, education, and medical training. Such analysis could help identify potential reasons behind variation in clinical practice. Besides the lack of information on the practitioners, the present study was limited by a paucity of research evidence in this area. We were also unable to retrieve consolidated data on both gender and age of patients registered in individual GP practices and therefore, assess the combined effect of sex and age distributions on test requesting activity. Furthermore, despite our attempt to maintain consistency by analysing laboratory test requests only from primary care medical practices that remained open throughout the study period, we acknowledge that some GPs in those practices could have been replaced, moved to alternative locations or quit direct patient care during the period of investigation. This could have an impact on the centre-associated requesting rates of laboratory tests. Finally, although the rationale of the study, research objectives, variables, inclusion criteria, outcome measures, and statistical methods to be used in our study were identified in advance to guide the analysis of the primary care use of laboratory tests in the WHSCT, the a-priori defined analysis plan was not published. We acknowledge that this could be perceived as a potential limitation of our study.

Conclusion

This study investigated the patterns and temporal changes in request rates across a range of frequently ordered laboratory tests. In addition, it explored potential factors of the inter-practice variability in the use of laboratory tests and found that differences in requesting activity appear unrelated to either demographic and socioeconomic characteristics of GP practices or clinical outcome indicators. Our results highlight the

need for further investigations to identify other potential factors that may account for the differences in test utilization between practitioners.

Acknowledgments

The authors would wish to thank Dr KongFatt Wong-Lin for helpful discussions and Graham Moore for computing and technical support.

References

1. Newman-Toker DE, McDonald KM, Meltzer DO, et al. How much diagnostic safety can we afford, and how should we decide? A health economics perspective. *BMJ Qual Saf.* 2013; 22: ii11-ii20.
2. Karakusevic S, Edwards N, Lewis R, et al. *The future of pathology services.* Nuffield Trust. 2016.
3. Smellie WSA. Demand management and test request rationalization. *Ann Clin Biochem* 2012; 49(4): 323-36.
4. Bareford D, Hayling A. Inappropriate use of laboratory services: long term combined approach to modify request patterns. *BMJ* 1990; 301(6764):1305-1307.
5. An Independent Review for the Department of Health. Report of the Review of NHS Pathology Services in England. Chaired by Lord Carter of Coles. 2006.
6. Smellie WSA, Galloway MJ, Chinn D, et al. Is clinical practice variability the major reason for differences in pathology requesting patterns in general practice? *Journal of Clinical Pathology*. 2002;55(4):312-314.
7. Sá L, Teixeira AS, Tavares F et al. Diagnostic and laboratory test ordering in Northern Portuguese Primary Health Care: a cross-sectional study. *BMJ Open*. 2017;7(11):e018509.
8. Kwok J, Jones B. Unnecessary repeat requesting of tests: an audit in a government hospital immunology laboratory. *Journal of Clinical Pathology*. 2005;58(5):457-462.
9. Vaidya B, Ukoumunne OC, Shuttleworth J, et al. Variability in thyroid function test requests across general practices in south-west England. Qual Prim Care. 2013;21: 143-8.
10. O'Kane MJ, Casey L, Lynch PM, et al. Clinical outcome indicators, disease prevalence and test request variability in primary care. *Ann Clin Biochem* 2011; 48(2): 155-158.

11. Smellie W, Galloway M, Chinn D. Benchmarking general practice use of pathology services: a model for monitoring change. *Journal of Clinical Pathology*. 2000;53(6):476-480.
12. Weydert JA, Nobbs ND, Feld R, et al. A simple, focused, computerized query to detect overutilization of laboratory tests. *Arch Pathol Lab Med* 2005;129(9):1141-1143.
13. Wong ET, McCarron MM, Shaw ST. Ordering of laboratory tests in a teaching hospital: can it be improved? *JAMA* 1983;249(22):3076-3080.
14. Young DW. Improving laboratory usage: a review. *Postgrad Med J* 1988;64(750):283-289.
15. Epstein AM, McNeil BJ. Relationship of beliefs and behavior in test ordering. *Am J Med* 1986; 80:865–870.
16. van Diepen S, Bakal JA, Lin M et al. Variation in critical care unit admission rates and outcomes for patients with acute coronary syndromes or heart failure among high- and low-volume cardiac hospitals. *J Am Heart Assoc* 2015; 4:e001708.
17. Adam MA, Goffredo P, Youngwirth L et al. Same thyroid cancer, different national practice guidelines: When discordant American Thyroid Association and National Comprehensive Cancer Network surgery recommendations are associated with compromised patient outcome. *Surgery* 2016; 159:41–50.
18. Northern Ireland Statistics and Research Agency. *Review of the Statistical Classification and Delineation of Settlements*. 2015. Available at: http://www.nisra.gov.uk/archive/geography/review-of-the-statistical-classification-and-delineation-of-settlements-march-2015.pdf. Accessed: Aug 2018.
19. Northern Ireland Statistics and Research Agency. *Northern Ireland Multiple Deprivation Measures 2017*. 2017. Available at: https://www.nisra.gov.uk/publications/nimdm17-results. Accessed: Jul 2018.
20. Department of Health. Quality and outcomes framework (QOF) achievement data 2015/16. 2016. Available at: https://www.health-ni.gov.uk/publications/quality-and-outcomes-framework-qof-achievement-data-201516. Accessed: Aug 2018

21. Department for Environment, Food & Rural Affairs. *Rural Urban Classification.* 2016. Available at: https://www.gov.uk/government/collections/rural-urban-classification. Accessed: Aug 2018.
22. Sen A, Srivastava M. *Regression analysis: theory, methods, and applications.* Springer Science & Business Media; 2012.
23. Jekel JF, Katz DL, Elmore JG et al. *Epidemiology, biostatistics and preventive medicine.* Elsevier Health Sciences; 2007.
24. Conover WJ, Johnson ME, Johnson MM. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*. 1981;23(4):351-61.
25. Li WK, Stanford DA, Yu H. *Advances in Time Series Methods and Applications.* Springer; 2016.
26. Gibbons JD, Fielden JD. *Nonparametric measures of association*. Sage. 1993
27. Gupta S, Verma M, Gupta AK, et al. Are we using Thyroid Function Tests Appropriately? *Indian Journal of Clinical Biochemistry* 2011;26(2):178-181.
28. Mogyorosy, Z, Mogyorosy, G. Practice pattern and geographic variation in test ordering: a literature review. *Orv Hetil* 2006;147:25–31.
29. Larsson, A, Palmer, M, Hulten, G, et al. Large differences in laboratory utilisation between hospitals in Sweden. *Clin Chem Lab Med* 2000;38:383–9.
30. Selvin, E, Steffes, MW, Zhu, H, Glycated haemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. *N Engl J Med* 2010;362:800–11.
31. Miyakis, S, Karamanof, G, Liontis, M, Mountokalakis, TD. Factors contributing to inappropriate ordering of tests in an academic medical department and the effect of an educational feedback strategy. Postgrad Med J 2006;82:823–9
32. Beckett, GJ, Toft, AD. First-line thyroid function tests – TSH alone is not enough. *Clin Endocrinol* 2003;58:20–1.
33. Bucholc M, O'Kane M, Ashe S, et al. Prescriptive variability of drugs by general practitioners. *PloS One* 2018;13(2):e0189599.

34. Morrison J, Anderson MJ, Sutton M, Munoz-Arroyo R, McDonald S, Maxwell M, et al. Factors influencing variation in prescribing of antidepressants by general practices in Scotland. *Br J Gen Pract* 2009;59(559): e25-e31.

35. Holm M, Olesen F. Factors affecting prescription of psychotropic drugs in general practice. *Scandinavian journal of primary health care* 1988;6(3):169-73.

36. Bugter-Maessen AM, Winkens RA, Grol RP, et al. Factors predicting differences among general practitioners in test ordering behaviour and in the response to feedback on test requests. *Fam Pract.* 1996;13: 254–258.

37. Baker R, Falconer Smith J, Lambert PC. Randomised controlled trial of the effectiveness of feedback in improving test ordering in general practice. *Scand J Prim Health Care.* 2003;21(4):219-223.

38. Buntinx F, Knottnerus JA, Crebolder HF, et al. Reactions of doctors to various forms of feedback designed to improve the sampling quality of cervical smears. *Qual Assur Health Care* 1992;4(2):161-166.

39. Bunting PS, Van Walraven C. Effect of a controlled feedback intervention on laboratory test ordering by community physicians. *Clin Chem.* 2004; 50(2):321-326.

40. Showstack JA, Schroeder SA, Matsumoto MF. Changes in the use of medical technologies, 1972–1977: a study of 10 inpatient diagnoses. *N Engl J Med.*1982;306: 706–712.

41. Pysher TJ , Bach PR , Lowichik A , et al . Chemistry test ordering patterns after elimination of predefined multitest chemistry panels in a children's hospital. *Pediatr Dev Pathol* 1999; 2:446–53.

42. Martins CM , da Costa Teixeira AS , de Azevedo LF , et al . The effect of a test ordering software intervention on the prescription of unnecessary laboratory tests - a randomized controlled trial. *BMC Med Inform Decis Mak* 2017; 17:20.

43. Oxman AD, Thomson MA, Davis DA, Haynes RB. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *Can Med Assoc J* 1995; 153:1423-31.

44. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993; 342:1317-22.

Contributors: MB and MJO had the original idea for this study. SA led the data collection. MB designed the methodology, performed the analysis, and drafted the manuscript. MJO, CM, and LM contributed to the drafting and critical revision of the manuscript.

Funding: This project was supported by the EU's INTERREG VA Programme, managed by the Special EU Programmes Body (SEUPB). The views and opinions expressed in this paper do not necessarily reflect those of the European Commission or the Special EU Programmes Body (SEUPB).

Competing interests: None declared.

Ethics approval: Not applicable.

Data sharing statement: Data on the use of laboratory tests in primary care in the period 2011-2016 can be obtained from the Clinical Chemistry departments of the Western Health and Social Care Trust (WHSCT) (info.enquiry@westerntrust.hscni.net), subject to WHSCT approval. The statistical analysis plan was not published a priori.

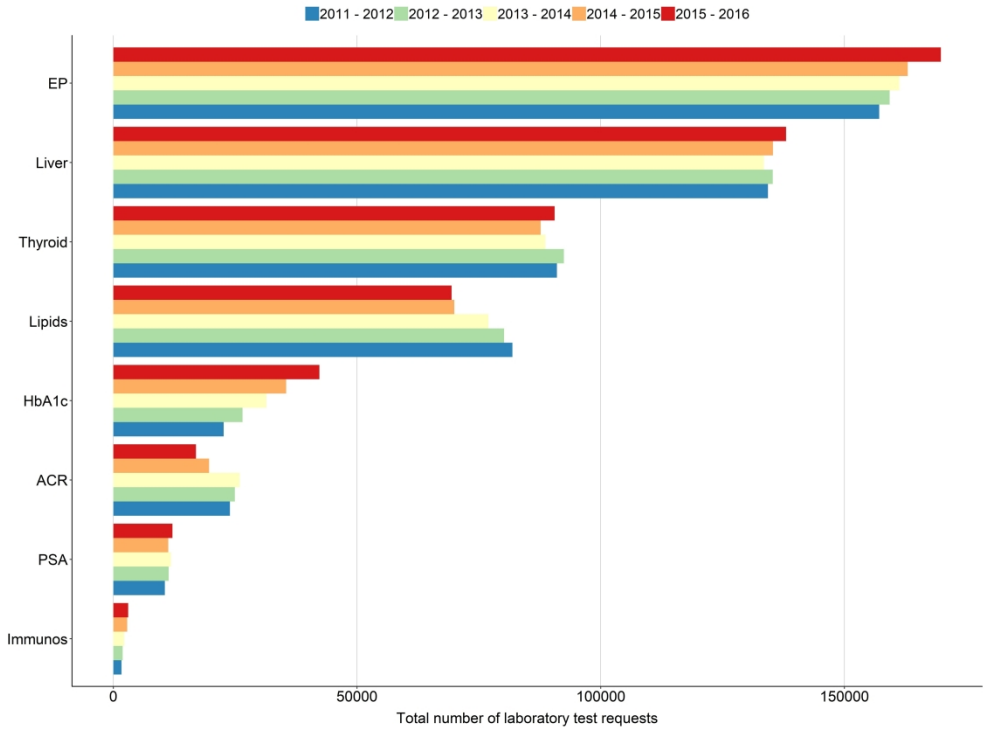
Figures

Figure 1. The total number of 8 considered clinical biochemistry tests/test groups ordered by 55 general practices during the period 1 Apr 2011 – 31 Mar 2016.

Figure 2. Temporal variability of the standardized laboratory test requests for (a) electrolyte profiles, (b) lipid profile, (c) prostate-specific antigen (PSA), (d) thyroid profiles, (e) liver profiles, (f) HbA_{1c}, (g) urine albumin/creatinine ratio (ACR), and (h) immunoglobulins (Immunos) for 55 considered general practices. Each data point (dot): a single practice. Solid, horizontal line inside the box: median. Lower and upper "hinges" of the boxplots: 1st and 3rd quartiles, respectively. Lower and upper extremes of whiskers: interval boundaries of the non-outliers (black dots). Data outside interval: outliers.

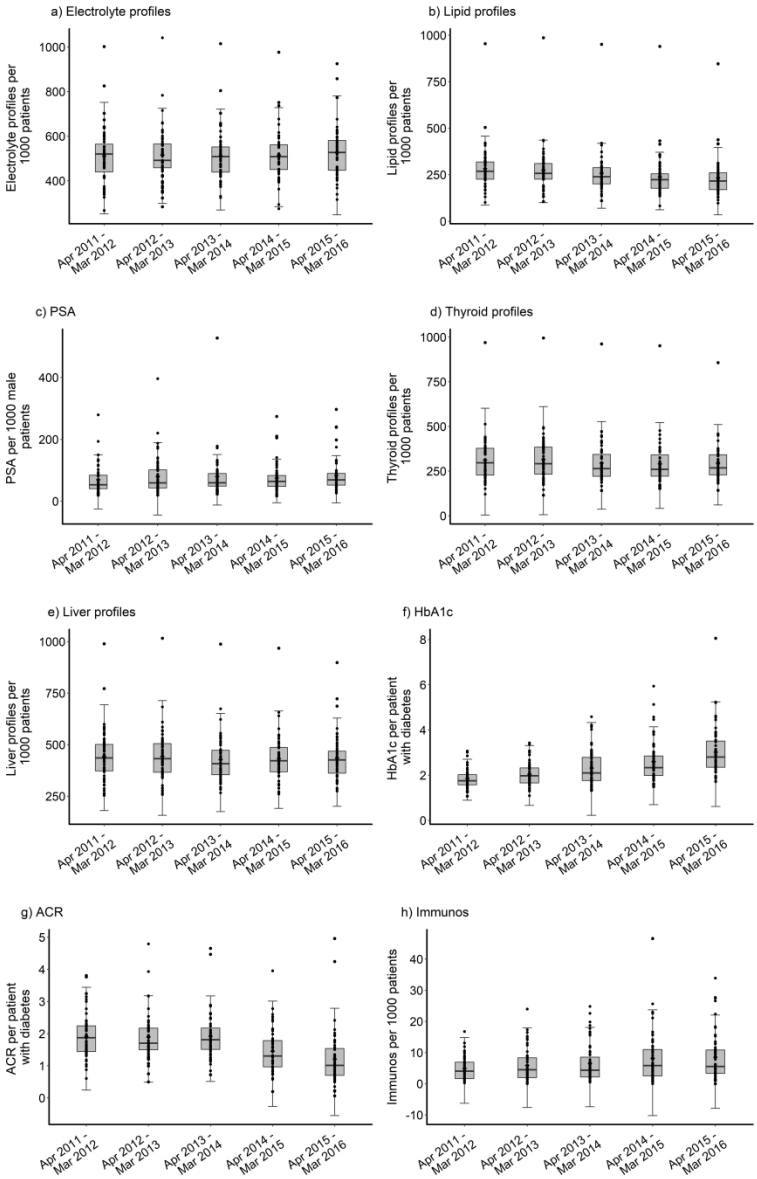
Figure 3. Trend lines for median (blue) and variance (red) of the standardized number of (a) electrolyte profiles, (b) lipid profile, (c) prostate-specific antigen (PSA), (d) thyroid profiles, (e) liver profiles, (f) HbA_{1c}, (g) urine albumin/creatinine ratio (ACR), and (h) immunoglobulins (Immunos).

Figure 4. The temporal median-variance relationship of the standardized number of laboratory test requests across the years for (a) electrolyte profiles (EP), (b) lipid profile (LP), (c) prostate-specific antigen (PSA), (d) thyroid profiles (TP), (e) liver profiles (LFT), (f) HbA_{1c}, (g) urine albumin/creatinine ratio (ACR), and (h) immunoglobulins (Immunos). Circle/triangle: rural/urban general practices.



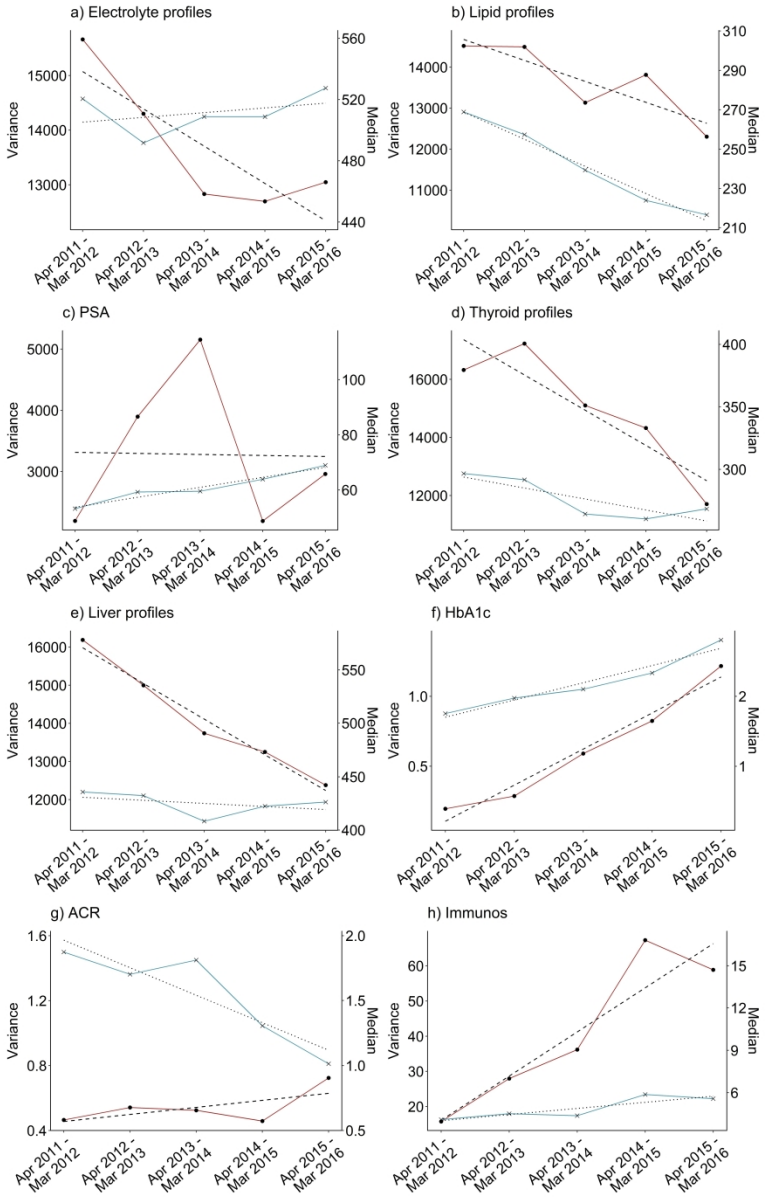
The total number of 8 considered clinical biochemistry tests/test groups ordered by 55 general practices during the period 1 Apr 2011 – 31 Mar 2016.

406x304mm (300 x 300 DPI)



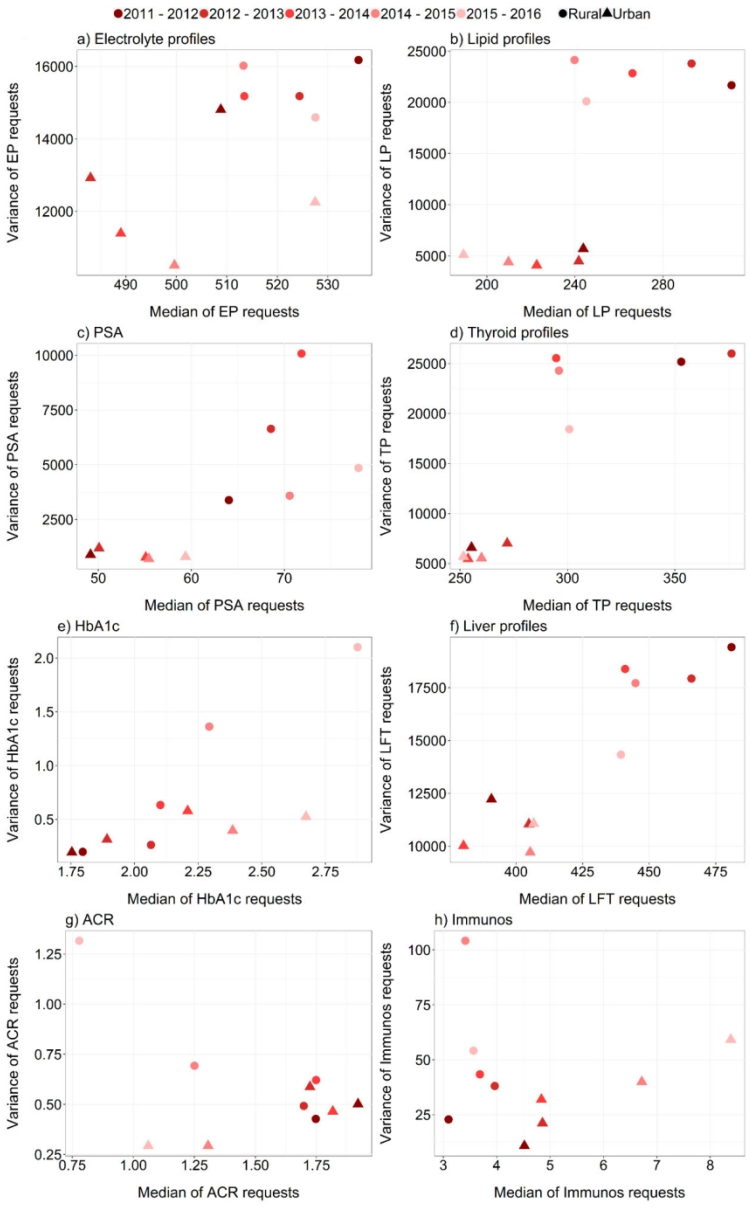
Temporal variability of the standardized laboratory test requests for (a) electrolyte profiles, (b) lipid profile, (c) prostate-specific antigen (PSA), (d) thyroid profiles, (e) liver profiles, (f) HbA1c, (g) urine albumin/creatinine ratio (ACR), and (h) immunoglobulins (Immunos) for 55 considered general practices. Each data point (dot): a single practice. Solid, horizontal line inside the box: median. Lower and upper "hinges" of the boxplots: 1st and 3rd quartiles, respectively. Lower and upper extremes of whiskers: interval boundaries of the non-outliers (black dots). Data outside interval: outliers.

355x558mm (300 x 300 DPI)



Trend lines for median (blue) and variance (red) of the standardized number of (a) electrolyte profiles, (b) lipid profile, (c) prostate-specific antigen (PSA), (d) thyroid profiles, (e) liver profiles, (f) HbA1c, (g) urine albumin/creatinine ratio (ACR), and (h) immunoglobulins (Immunos).

355x558mm (300 x 300 DPI)



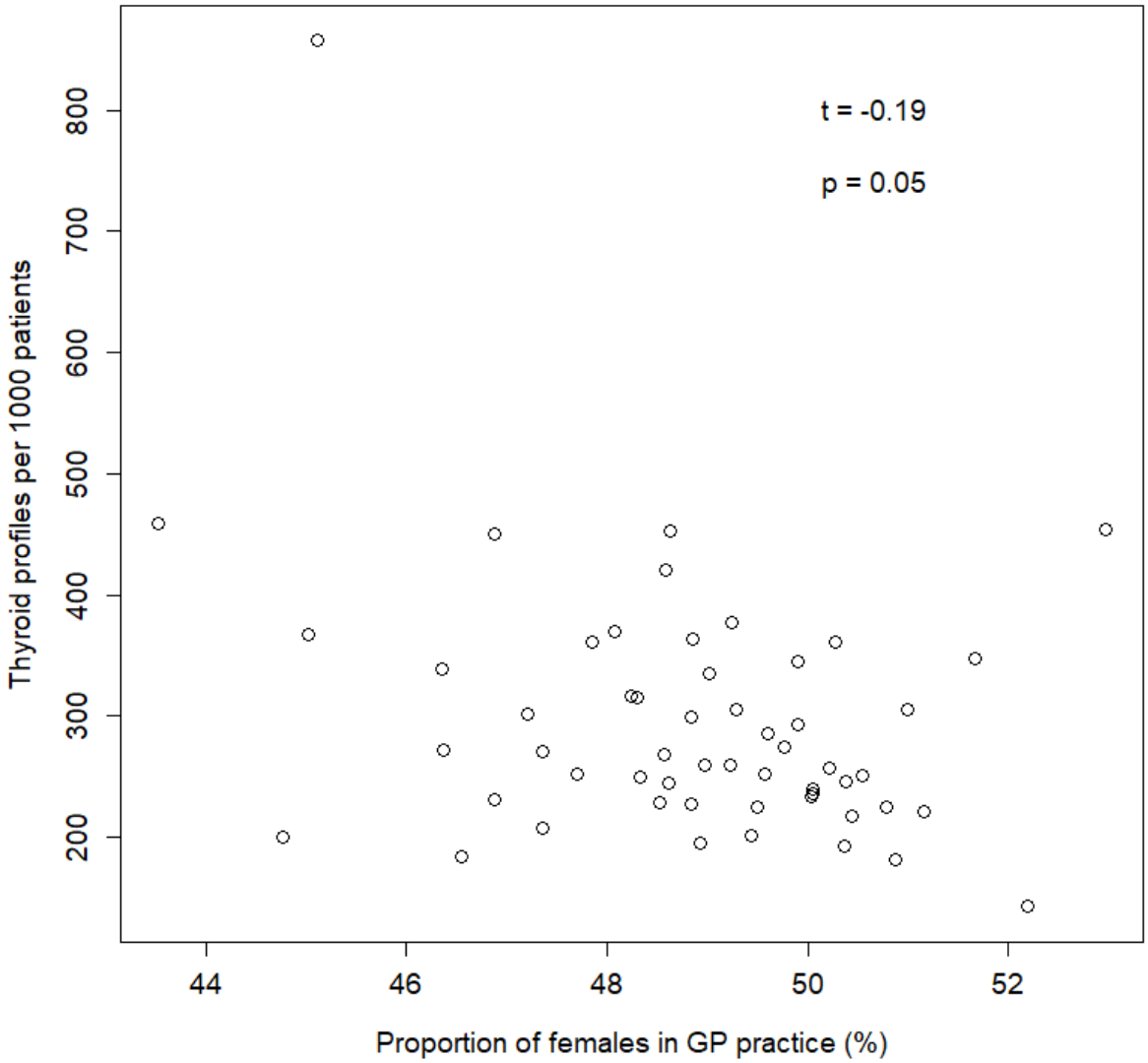
The temporal median-variance relationship of the standardized number of laboratory test requests across the years for (a) electrolyte profiles (EP), (b) lipid profile (LP), (c) prostate-specific antigen (PSA), (d) thyroid profiles (TP), (e) liver profiles (LFT), (f) HbA1c, (g) urine albumin/creatinine ratio (ACR), and (h) immunoglobulins (Immunos). Circle/triangle: rural/urban general practices.

105x169mm (300 x 300 DPI)

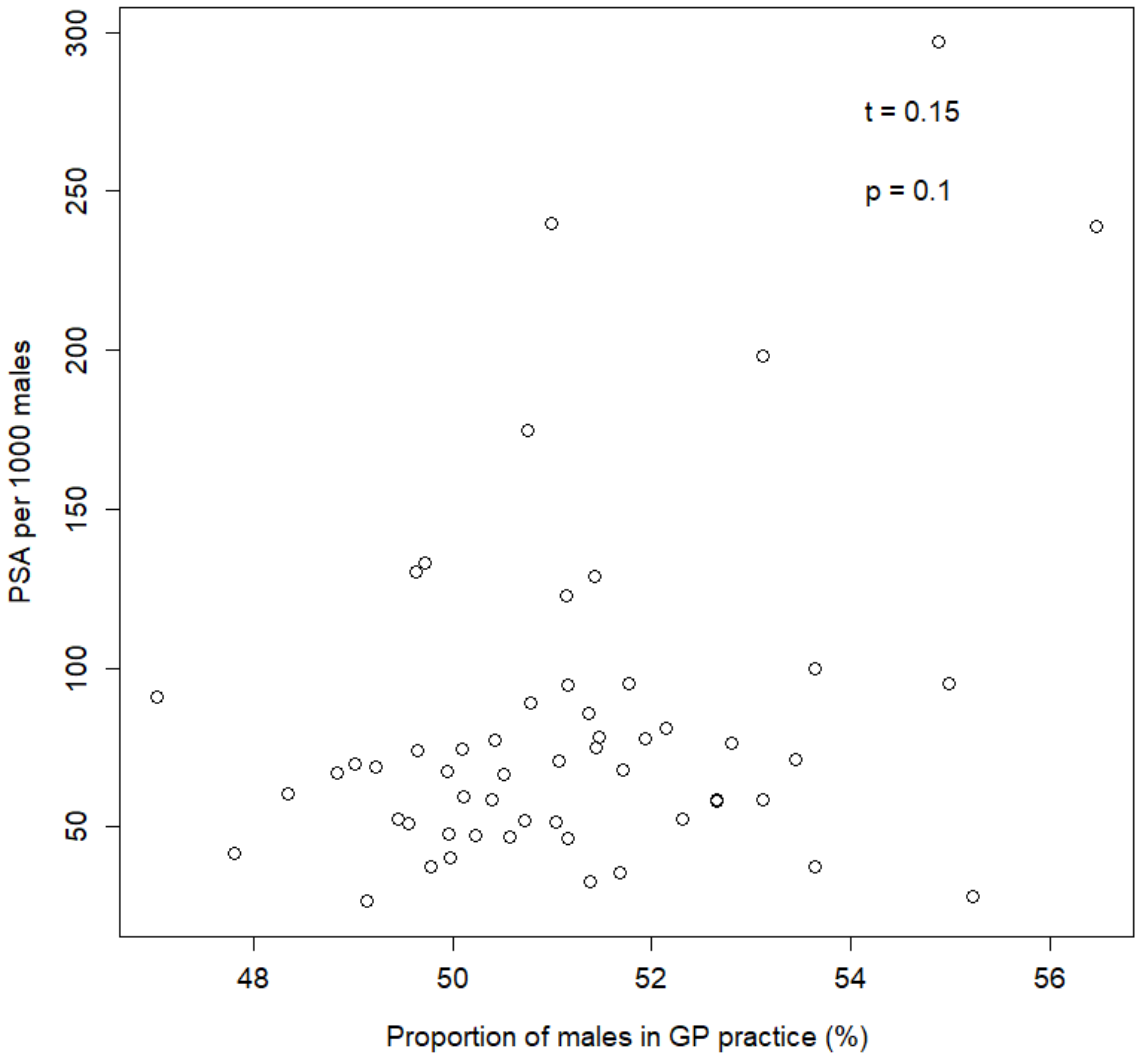
Supplementary Table S1. The relationship between test requesting rates and demographic characteristics of patient population. Kendall's correlation coefficient tau: τ. p-value associated with τ: $p_{Kendall}$. Bonferroni-corrected significance cut-off: $\alpha = p_{fwe}/n < 0.0125$ where $p_{fwe} < 0.05$ is the family wise error probability and $n = 4$ is the number of tested associations.

Age group	Electrolyte profiles	Lipid profiles	Thyroid profiles	Liver profiles
≥ 65 years old				
τ	0.21	0.23	0.08	0.20
$p_{Kendall}$	0.04	0.02	0.4	0.05

Supplementary Figure S1. The relationship between the percentage of females in individual GP practices and requesting rates for thyroid profiles.



Supplementary Figure S2. The relationship between the percentage of males in individual GP practices and requesting rates for PSA.



Supplementary Table S2. The relationship between test requesting rates and socioeconomic factors. Kendall's correlation coefficient tau: τ; p-value associated with τ: $p_{Kendall}$. Bonferroni-corrected significance cut-off: $\alpha = p_{fwe}/n < 0.003125$ where p_{fwe} 0.05 is the family wise error probability and $n = 16$ is the number of tested associations.

	Electrolyte profiles	Lipid profiles	PSA	Thyroid profiles	HbA1c	Liver profiles	ACR	Immunoglobulins
Multiple Deprivation Measure								
τ	0.034	0.116	0.180	0.119	0.012	0.083	-0.094	-0.167
$p_{Kendall}$	0.7	0.2	0.05	0.2	0.9	0.4	0.3	0.08
Health Deprivation & Disability								
τ	0.124	0.256	0.209	0.239	0.062	0.185	-0.021	-0.156
$p_{Kendall}$	0.2	0.007	0.013	0.01	0.5	0.05	0.8	0.1

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	5-6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6-7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-8
Bias	9	Describe any efforts to address potential sources of bias	6-7,20
Study size	10	Explain how the study size was arrived at	6-7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6-8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8-9
		(*b*) Describe any methods used to examine subgroups and interactions	9
		(*c*) Explain how missing data were addressed	Not applicable
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	Not applicable
		(*e*) Describe any sensitivity analyses	Not applicable
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Not applicable
		(b) Give reasons for non-participation at each stage	Not applicable
		(c) Consider use of a flow diagram	Not applicable
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Not applicable
		(b) Indicate number of participants with missing data for each variable of interest	Not applicable
Outcome data	15*	Report numbers of outcome events or summary measures	Not applicable
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Not applicable
		(*b*) Report category boundaries when continuous variables were categorized	16
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	10-16
Discussion			
Key results	18	Summarise key results with reference to study objectives17	17-18
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	20
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	19
Generalisability	21	Discuss the generalisability (external validity) of the study results	17-18
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	26

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.