BMC Medicine

Practice variation in the use of tests in UK primary care: a retrospective analysis of 16 million tests performed over 3.3 million patient years in 2015/16.

--Manuscript Draft--

Manuscript Number:	BMED-D-18-01116R1
Full Title:	Practice variation in the use of tests in UK primary care: a retrospective analysis of 16 million tests performed over 3.3 million patient years in 2015/16.
Article Type:	Research article
Section/Category:	Health Services Research
Funding Information:	School for Social Care Research (386) — Dr Jack William O'Sullivan
Abstract:	Background The UK's National Health Service (NHS) is currently subject to unprecedented financial strain. The identification of unnecessary healthcare resource use has been suggested to reduce spending. However, there is little very research quantifying wasteful test use, despite the £3 billion annual expenditure. Geographical variation has been suggested as one metric in which to quantify inappropriate use. We set out to identify tests ordered from UK primary care that are subject to the greatest between-practice variation in their use. Methods We used data from 444 general practices within the Clinical Practice Research Datalink to calculate a co-efficient of variation (CoV) for the ordering of 44 specific tests from UK general practices. The co-efficient of variation was calculated after adjusting for differences between practice populations. We also determined the tests that had both a higher-than-average CoV and a higher-than-average rate of use. Results In total, 16,496,218 tests were ordered for 4,078,091 patients over 3,311,050 person-years from April 1st 2015 to March 31st 2016. The tests subject to the greatest variation was drug monitoring 158% (95%CI: 153% to 163%), Urine Microalbumin (52% (95%CI: 49.9% to 53.2%)), Pelvic CT (51% (95%CI: 50% to 53%)) and Pap smear (49% (95%CI: 48% to 51%). Seven tests were classified as high variability and high rate (Clotting, Vitamin D, Urine Albumin, Prostate Specific Antigen (PSA), Bone profile, Urine MCS and C-reactive Protein (CRP)). Conclusions There are wide variations in the use of common tests, which is unlikely to be explained by clinical indications. Since £3 billion annually are spent on tests this represents considerable variation in the use of resources and inefficient management in the NHS. Our results can be of value to policy makers, researchers, patients, and clinicians as the NHS strives towards identifying overuse and underuse of tests. Funding National Institute for Health Research School of Primary Care Research (Award Number 386)
Corresponding Author:	Jack William O'Sullivan, MBBS, DPhil University of Oxford Oxford, Oxfordshire UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Oxford
Corresponding Author's Secondary Institution:	
First Author:	Jack William O'Sullivan, MBBS, DPhil
First Author Secondary Information:	
Order of Authors:	Jack William O'Sullivan, MBBS, DPhil

	Sarah Stevens, DPhil
	Jason Oke, DPhil
	Richard Hobbs, FMedSci
	Chris Salisbury, MD
	Paul Little, FMedSci
	Ben Goldacre, MRCPsych
	Clare Bankhead, DPhil
	Jeffrey Aronson, DPhil
	Carl Heneghan, DPhil
	Rafael Perera, DPhil
Order of Authors Secondary Information:	
Response to Reviewers:	Dear Dr Lopez Munoz, Thank you for the opportunity to revise our paper. We would also like to thank the peer reviewers for their comments. Reviewer #1: This is original, important and methodologically rigorous enough to merit publication as it stands. Response: Thank you for the comment, no response required. Reviewer #2: The topic is important, the methods are relevant and the conclusions interesting. i only have minor comments: - The selection of the 44 tests is justified by "common use." It might be useful to know more: How common? How many tests were available in total? Response: The CPRD contains codes for around 550 different tests. As the reviewer points out, one of the criteria we used to select our 44 specific tests was their frequency of use. We describe the process we undertook to select these tests in the supplementary file (Page 3). Briefly, we obtained data directly from Oxford University Hospital (which processes all NHS primary care test requests) and determined the two most frequent laboratory and imaging tests ordered from Oxfordshire primary care. To supplement this, we searched the Quality Outcomes Framework (QOF), National Institute of Health and Clinical Excellence (NICE), Choosing Wisely and NICE Do Not Do guidelines. We identified all the tests mentioned in these guidelines and included these in our analysis as well. Furthermore, during preliminary data cleaning it became apparent that some tests are ordered as one test, but then return numerous results. For instance, a full blood count is commonly ordered as one test, but its results are stratified (i.e. haemoglobin, Mean corpuscular volume (MCV) etc) and thus can be coded separately. For these tests we grouped their codes and counted them as one test. These tests were also included. - There is little discussion of whether the assumptions behind the Poisson model are satisfied (What is the man compared to the variance?) Response: We agree that it would be advantageous if we were more clear about our use of the Poisson model. The main focus of our use of a Poisson model was for adjustment of suspected modifiers (Age, Sex and IMD) in order to create an adjusted measure of variation. As our outcome of interest is count data within a set time period a Poisson model is the natural approach. To expand, we constructed the Poisson model to adjust for differences in patient demographics between general practices. The Poisson models allowed us to determine the rate of test use (and thus the co-efficient of variation) once differences in patient demographics (sex, age and deprivation (IMD)) were accounted for. It was not the aim of the Poisson model (and the paper more broadly) to determine if one of our patient demographics was a better predictor of test use than another, i.e. we did not

	aim to determine if age was a better predictor of test use than IMD. We expected there would be variation in test use between general practices (which there was), and we expected that our covariates (age, sex and IMD) would help account for some, but not all, of this variation (which it did - see figure 2 and appendix figure 1 (supplementary file)). We were not concerned by residual variation that persisted in our results from the Poisson model ("overdispersion"). This residual variation is what we suggest as the variation in test use once practice differences in patient age, sex and IMD have been accounted for. As is suggested in the literature (1), a model that entirely accounts for overdispersion (and thus entirely satisfies the model assumptions) when examining geographical variation in healthcare use would be inappropriate, as, at the very least, there will also be natural sample variation and human variation. Therefore, as one might expect, the satisfaction of Poisson model assumptions varied among the 44 models (one for each test). However, given the expected overdispersion and our desire to rank tests based on their overdispersion, the degree in which these assumptions were met were not central to the results and implications to this paper. We did, however, test the impact of this by construction of a quasipoisson model. A Poisson model assumes a dispersion parameter of 1 (the mean is equal to the variance), while a quasipoisson model does not assume a dispersion parameter. When we rerun our analysis using a quasipoisson model, our calculated adjusted rates and adjusted CoV were unchanged (the standard errors changed, but they are not important for what we are presenting in this paper). Thus the results of this paper are unchanged regardless of whether this assumption (mean = variance) is met or not. We have tried to include the above considerations with the following additional paragraphs in the manuscript. In the methods we added: "We constructed Poisson models to adjust for differences in patient demographics (age, sex and IMD) between general practices. We did not construct Poisson models to compare the predictive ability of patient demographics of rates of test use." In the discussion we added: "We used the conventional statistical analysis for count data; a Poisson regression model. However, we used the outputs of this model in a less conventional manner. The aim of this paper was to determine which tests were subject to the most between-variation practice in their use once patient demographics between practices had been accounted for. We did not use the Poisson models to determine and compare the predictive ability of our covariates (patient demographics). As is expected when analysing health care data (20), the model accounted for some, but not all, of the variation in test use. This residual variation – "overdispersion" – represents the variation in test use once patient demographics between practices had been accounted for. We ranked tests by their residual variation; variation in use that persisted despite adjustment of patient demographic differences between practices." - When the analysis is done on the 444 practices, one is likely to find more variation since there will be some that focus on special groups (see the extreme outlier - non illicit drug testing - in one figure) and this is not fully captured in the Poisson model which only has three variables. A different approach would be to group the practices into regions and examine differences between regions. Assuming patients in regions are more comparable than patients in single practices, this would be a better way of identifying unwarranted variation and not just variation that is caused by patient differences. However, this would be a new paper and the paper is interesting as it is. Response: This is an interesting idea and we agree it would be of value to explore this concept in a new paper. Although, if one was to follow the approach suggested - the aggregation of practices into regions and examine variation between regions - one may risk obfuscating true, unwarranted variation. As the reviewer points out, it is highly likely that more variation exists between general practices (within regions) than between regions. Further, in the database that we used (the CPRD), regional breakdown is very broad. For instance, Wales, Scotland and Northern Ireland are considered only one region respectively. England is broken into 10 regions. Given the likely heterogeneity of data (practices, patient demographics etc) when accumulated into regions, the results of an aggregated regional analysis, would likely obfuscate true, unwarranted variation. For context, there are more than 200 clinical commissioning groups (CCGs) in the UK.

Further to this point, we have previously done a paper (https://www.nature.com/articles/s41598-018-23263-z) that shows that practices within the same CCG (Oxfordshire) can vary quite substantially; both in terms of patient demographics and also the number of tests they ordered (per patient). For instance, among the 69 practices we examined in the Oxfordshire CCG, the IMD varied from 3 to 10 (1 is the most deprived, 10 the most). Thus, although this is an interesting idea and a potential idea for a follow up paper, we currently favour using practices, rather than regions are the unit of analysis. Nevertheless, we have added the following paragraph to the future research section of the discussion to address the reviewer's valid suggestion (note it also is a response to a comment from reviewer 3):

"It would be advantageous for future studies to investigate variation using a different unit of an analysis. We chose to investigate variation at a practice level, however future studies could investigate variation at a patient-level or at a regional level. We chose to investigate at a practice level as previous literature suggests practice level factors contribute substantial to healthcare variation (2,3). Differences in disease prevalence, cultural attitude to tests and risks, local key-opinion leaders, resource availability, local policy and guidelines, and service configurations have all been suggested as practice-level contributors to variation.
A similar analysis aggregated at a regional, rather than practice, level may provide further insight into unwarranted variation. It is plausible that our analysis at a practice level may be too sensitive to variation in disease prevalence, this may in part explain non-illicit drug testing as an outlier. However, the aggregation of data at a regional level may obfuscate true, unwarranted variation. Furthermore, the CPRD only allows practices to be identified at a broad regional level (e.g. within Wales). Conversely, future research that analyses data at an individual patient level may provide more nuanced insight into variation, but risks being overly sensitive; making the distinction between warranted and unwarranted variation more difficult. Nevertheless, we would welcome any further studies using the aforementioned analyses."

- A more systematic explanation of factors that could explan the patter (incentives that differ? culture? geographic location?) would be useful, but also difficult and maybe the subject of another article.

Response: We agree with the reviewer that it is plausible that the variation in test use we present could be further explained by other factors. It is plausible that some of these factors are differing incentives, although this is less likely in the UK as general practice activity is incentivised under a national scheme (the quality outcomes framework (QOF)). It is also plausible that cultural differences between different geographical locations may also explain some of the residual variation. Unfortunately the database we used (CPRD) does not contain any further practice (or patient) covariates that we could add to our model. It was also not possible for us to cross reference our CPRD data into another UK datasource because CPRD data is anonymised at both a patient and practice level.

Reviewer #3: Geographical variations in practice are an under-researched but really important area so it is great to see this draft. Some more care needs to be taken in some of your assumptions, though.

Response: Thank you, no response required.

Page 4, lines 13-16. As primary care accounts for most health care (90% of all UK National Health Service (NHS) care [7], 55% in the USA [8]), it is the ideal target for reducing overuse. Primary care certainly accounts for the majority of health care encounters. However, it only accounts for around 7-8% of total NHS spend so this sentence seems a bit naive.

Response: This is a good point by the reviewer and we have amended the sentence to make it more circumspect. It now reads:
"To help reduce costs, the identification of unnecessary care has become a focus of governments and healthcare funders around the world (4). Primary care accounts for most health care (90% of all UK National Health Service (NHS) care (5), 55% in the USA (6)) and serves a gatekeeper function in the UK; tests often have knock-on

consequences in both primary and secondary care. As such, the identification of wasteful resource use in primary care has implications the entire healthcare system."

Page 4, line 23. there is little research into variation in the use of tests by general practitioners (GP). Technically not really. Around a decade ago, there was masses of stuff on variation in QoF indicators, many of which related to the use of tests.

Response: We are not aware of any research papers that report between-practice variation in UK primary care test use. We are aware of longitudinal studies assessing activity of QOF compared with non-QOF activity (7,8) and one study that looks at geographical variation in primary care test use (9). We are also aware of the UK Atlases of variation, one of which addresses variation in test use. The major difference between the atlases of variation and our study is the atlases identify regions that vary from the national average, whereas our study identifies which tests are subject to the greatest variation in their use. The atlases also use regions as there unit of analysis (we use practices). If the reviewer has access to any other previous research that would be beneficial for our study, we would be very grateful.

Page 5. My major point of confusion lies in the statistical methods you have chosen (from line 30). I don't understand why, if you had age and sex information at the individual level, you didn't use this to get a more nuanced insight into variation in test rates. There may be a good reason but perhaps you could explain why (given the claims made about the quality of the individual level data). The description of the Poisson regression model is, to be honest, overly brief and left me thinking that your adjusted rates were based on the model predicted rate (given the practices age, sex and IMD decile) rather than the actual rate. This doesn't seem appropriate but if this is what was actually done, you need to say why. If it's not what was done, you need to be clearer!

Response:

There are two points to respond to here:
1.The description of the statistical analysis and Poisson regression model
2.Justification of the statistical model we chose

Description of the statistical analysis and Poisson regression model
We calculated and present both the unadjusted coefficient of variation and the adjusted coefficient of variation. We use the term "unadjusted" to refer to the raw, crude coefficient of variation (CoV). We calculated the adjusted and unadjusted CoV by the following steps:
1.Unadjusted (crude) rate: The total number of tests ordered for each 44 specific tests from each general practice divided by the person-years for each general practice (For the period April 1st 2015 to 31st March 2016). I.e. We calculated the unadjusted rate of CRP test use from each 444 general practices.
2.Unadjusted (crude) CoV: We then calculated the mean unadjusted rate of use for each specific test across all 444 general practices, we also calculated the corresponding standard deviation. We used these two numbers to calculate the unadjusted CoV (standard deviation/mean x 100). We present the unadjusted CoV for each 44 tests in the supplementary (Appendix Table 1 and Appendix Figure 1).
3.Adjusted rate: We constructed a Poisson model with IMD, age and sex covariates to calculated the adjusted number of tests ordered for each specific test from each general practice. We then divided this adjusted number of tests by the person-years for each general practice - this is our adjusted rate. Thus, the Poisson model allowed us to adjust the numerator from our "unadjusted rate" calculation.
4.Adjusted CoV: We then followed the same process described in "unadjusted CoV" to calculate an adjusted CoV, i.e. we calculated the mean adjusted rate of use for each specific test across all 444 general practices, we also calculated the corresponding standard deviation. We then used these two numbers to calculate the adjusted CoV (standard deviation/mean x 100). We present the adjusted CoV in the main paper (Figure 1).

Thus, to answer the reviewer's question, we present both the adjusted and unadjusted rate and the respective coefficient of variation (we presume the reviewer's use of "actual rate" is what we referred to as the unadjusted). We chose to present the

	adjusted rate in the main manuscript (we present both in the supplementary file), because the adjusted rate controls for differences in patient demographics between general practices. Our adjusted rate and adjusted CoV calculations control for the differences in age, sex and deprivation (which broadly includes co-morbidities) between practices. If these differences between practices were not controlled for, variation between practices may appear larger than it actually is (see Appendix table 1 and Appendix figure 1 in the supplementary file, which shows the larger variation in the unadjusted estimates compared with the adjusted estimates). To make the paper clearer, we have completely rewritten the “Statistical Analysis” section in the Methods section in line with the above response. We have also given a more detailed description of the Poisson model. We now state in the methods: “we constructed a generalised linear model with Poisson errors to estimate the number of tests ordered from each general practice adjusted for practice differences in patient age, sex and deprivation. We constructed 44 Poisson models for each test. The age covariate represents the median age of each general practice, the sex covariate represents the proportion of female patients in each practice and the deprivation covariate represents the practice-level Index of Multiple Deprivation (IMD) deciles. We constructed Poisson models to adjust for differences in patient demographics (age, sex and IMD) between general practices. We did not construct Poisson models to compare the predictive ability of patient demographics on the rates of test use.” We also hope the second half of our response to this comment helps clarify our use of the Poisson model. Justification of the statistical model We agree with the reviewer that it is necessary to be clearer about our justification of the use of Poisson models. We constructed Poisson models as they are the appropriate model when the analysis concerns count data within a set period. We used the Poisson model to adjust for differences in patient demographics between general practices. The Poisson models allowed us to determine the rate of test use (and thus the co-efficient of variation) once differences in patient demographics (sex, age and deprivation (IMD)) were accounted for. It was not the aim of the Poisson model (and the paper more broadly) to determine if one of our patient demographics was a better predictor of test use than another, i.e. we did not aim to determine if age was a better predictor of test use than IMD. We decided to aggregate data at a practice rather than patient level for three reasons: 1. We did have age and sex data at an individual level, but we only had IMD at a practice level. 2. We were interested in investigating variation between practices, rather than variation between patients. We thought it would be more likely that unwarranted variation would exist between practices, compared with variation between patients. We believe this because previous literature suggests that variation may be explained by differences in disease prevalence, cultural attitude to tests and risks, local key-opinion leaders, resource availability, local policy and guidelines, and service configurations [2]. These factors are more relevant at a practice-level, rather than patient-level.3. We followed an a priori protocol, indicating our desire to analyse at a practice level. Nevertheless, to address this valid point: we added the following paragraph to the discussion (note that this paragraph also responds to a comment from reviewer 2): “It would be advantageous for future studies to investigate variation using a different unit of an analysis. We chose to investigate variation at a practice level, however future studies could investigate variation at a patient-level or at a regional level. We chose to investigate at a practice level as previous literature suggest practice level factors contribute substantial to healthcare variation (2,3). Differences in disease prevalence, cultural attitude to tests and risks, local key-opinion leaders, resource availability, local policy and guidelines, and service configurations have all been suggested as practice-level contributors to variation. A similar analysis aggregated at a regional, rather than practice, level may provide further insight into unwarranted variation. It is plausible that our analysis at a practice level may be too sensitive to variation in disease prevalence, this may in part explain non-illicit drug testing as an outlier. However, the aggregation of data at a regional level may obfuscate true, unwarranted variation. Furthermore, the CPRD only allows

practices to be identified at a broad regional level (e.g. within Wales). Conversely, future research that analyses data at an individual patient level may provide more nuanced insight into variation, but risks being overly sensitive; making the distinction between warranted and unwarranted variation more difficult. Nevertheless, we would welcome any further studies using the aforementioned analyses."

References

1. Spiegelhalter DJ. Handling over-dispersion of performance indicators. Qual Saf Heal Care [Internet]. 2005;14(5):347–51. Available from: http://qualitysafety.bmj.com/lookup/doi/10.1136/qshc.2005.013755
2. Appleby J, Raleigh V, Frosini F, Bevan G, Gao H, Lyscom T. Variations in health care: The good, the bad and the inexplicable. 2011.
3. Wennberg JE. Time to tackle unwarranted variations in practice. BMJ. 2011;342:d1513.
4. Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, Heath I, et al. Evidence for overuse of medical services around the world. Lancet [Internet]. Elsevier Ltd; 2017;6736(16):1–13. Available from: http://dx.doi.org/10.1016/S0140-6736(16)32585-5%5Cnhttp://linkinghub.elsevier.com/retrieve/pii/S0140673616325855
5. Hobbs FDR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. Lancet [Internet]. Elsevier; 2016 Jun [cited 2016 Aug 19];387(10035):2323–30. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0140673616006206
6. Centers for Disease Control and Prevention, National Center for Health Statistics. National Ambulatory Medical Care Survey: 2012 Summary Tables. 2012;5. Available from: http://www.cdc.gov/nchs/data/ahcd/namcs_summary/2010_namcs_web_tables.pdf
7. Doran T, Kontopantelis E, Valderas JM, Campbell S, Roland M, Salisbury C, et al. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. BMJ [Internet]. 2011 [cited 2016 Aug 27];342:d3590. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21712336
8. Ryan AM, Krinsky S, Kontopantelis E, Doran T. Long-term evidence for the effect of pay-for-performance in primary care on mortality in the UK: a population study. Lancet (London, England) [Internet]. 2016 Jul 16 [cited 2016 Aug 23];388(10041):268–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27207746
9. Busby J, Schroeder K, Woltersdorf W, Sterne JAC, Ben-Shlomo Y, Hay A, et al. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices: Using routine data to identify research priorities. Br J Gen Pract. 2013;63(609):256–66.

Practice variation in the use of tests in UK primary care: a retrospective analysis of 16 million tests performed over 3.3 million patient years in 2015/16.

Jack W. O'Sullivan,[1,2] Sarah Stevens,[2] Jason Oke,[2] FD Richard Hobbs,[2] Chris Salisbury,[3] Paul Little,[4] Ben Goldacre,[1,2] Clare Bankhead,[1,2] Jeffrey K. Aronson,[1,2] Carl Heneghan, [1,2] and Rafael Perera [1,2].

[1]Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Sciences, University of Oxford, OX2 6GG, UK

[2]Nuffield Department of Primary Care Health Sciences, University of Oxford, OX2 6GG, UK

[3]Centre for Academic Primary Care, Department of Population Health Sciences, Bristol Medical School, University of Bristol, BS8 2PS, UK

[4]Primary Care and Population Sciences, University of Southampton, Southampton, SO17 1BJ, UK

Jack W. O'Sullivan, Clinical Researcher and DPhil Candidate, jack.osullivan@phc.ox.ac.uk

Sarah Stevens, Senior Statistician, sarah.stevens@phc.ox.ac.uk

Jason Oke, Senior Statistician, jason.oke@phc.ox.ac.uk

FD Richard Hobbs, Nuffield Professor of Primary Care Health Sciences, richard.hobbs@phc.ox.ac.uk

Chris Salisbury, Professor of Primary Health Care, c.salisbury@bristol.ac.uk

Paul Little, Professor of Primary Care Research, P.Little@soton.ac.uk

Ben Goldacre, Senior Clinical Research Fellow, ben.goldacre@phc.ox.ac.uk

Clare Bankhead, Associate Professor of Primary Care, clare.bankhead@phc.ox.ac.uk

Jeffrey K. Aronson, Clinical Pharmacologist, jeffrey.aronson@phc.ox.ac.uk

Carl Heneghan, Professor of Evidence-Based Medicine, carl.heneghan@phc.ox.ac.uk

Rafael Perera, Professor of Medical Statistics, Rafael.perera@phc.ox.ac.uk

Correspondence to: Dr Jack W O'Sullivan

Centre for Evidence-Based Medicine

Nuffield Department of Primary Care Health Sciences

Radcliffe Observatory Quarter, Oxford, OX2 6GG

jack.osullivan@phc.ox.ac.uk

01865 289300

Manuscript word count: 2,751

Number of references: 35

Practice variation in the use of tests in UK primary care: a retrospective analysis of 16 million tests performed over 3.3 million patient years in 2015/16.

Abstract

Background

The UK's National Health Service (NHS) is currently subject to unprecedented financial strain. The identification of unnecessary healthcare resource use has been suggested to reduce spending. However, there is little very research quantifying wasteful test use, despite the £3 billion annual expenditure. Geographical variation has been suggested as one metric in which to quantify inappropriate use. We set out to identify tests ordered from UK primary care that are subject to the greatest between-practice variation in their use.

Methods

We used data from 444 general practices within the Clinical Practice Research Datalink to calculate a co-efficient of variation (CoV) for the ordering of 44 specific tests from UK general practices. The co-efficient of variation was calculated after adjusting for differences between practice populations. We also determined the tests that had both a higher-than-average CoV and a higher-than-average rate of use.

Results

In total, 16,496,218 tests were ordered for 4,078,091 patients over 3,311,050 person-years from April 1st 2015 to March 31st 2016. The tests subject to the greatest variation was drug monitoring 158% (95%CI: 153% to 163%), Urine Microalbumin (52% (95%CI: 49.9% to 53.2%)), Pelvic CT (51% (95%CI: 50% to 53%)) and Pap smear (49% (95%CI: 48% to 51%). Seven tests were classified as high variability and high rate (Clotting, Vitamin D, Urine Albumin, Prostate Specific Antigen (PSA), Bone profile, Urine MCS and C-reactive Protein (CRP)).

Conclusions

There are wide variations in the use of common tests, which is unlikely to be explained by clinical indications. Since £3 billion annually are spent on tests this represents considerable variation in the use of resources and inefficient management in the NHS. Our results can be of value to policy makers, researchers, patients, and clinicians as the NHS strives towards identifying overuse and underuse of tests.

Funding

National Institute for Health Research School of Primary Care Research (Award Number 386)

Abstract word count: 264

Keywords: Overuse, health policy, primary care, general practice, test use, imaging.

Introduction

Healthcare systems around the world are struggling to remain fiscally sustainable (1–3). With increases in spending, healthcare systems are faced with a mismatch between funding and expenditure (4,5).

To help reduce costs, the identification of unnecessary care has become a focus of governments and healthcare funders around the world (6). Primary care accounts for most health care (90% of all UK National Health Service (NHS) care (7), 55% in the USA (8)) and serves a gatekeeper function in the UK; tests often have knock-on consequences in both primary and secondary care. As such, the identification of wasteful resource use in primary care has implications the entire healthcare system.

Previous research has suggested that when there is strong evidence and a professional consensus that an intervention is effective, there tends to be almost no variation in practice (9,10). Conversely, variation in the use of resources has been used to highlight possible overuse or underuse (10,11).

Despite its contribution to care and expenditure (12), there is little research into variation in the use of tests by general practitioners (GP). UK GPs are thought to spend be more than £3 billion annually on tests (12). One study has explored variation in GP test use, but it focused on a few specific tests in a relatively small population (13). We set out to identify which tests are subject to the greatest between practice variation in their use.

Methods

Study population

We obtained electronic health record data from patients registered with general practices contributing to the Clinical Practice Research Datalink (CPRD) during April 1st 2015 to 31st March 2016. The CPRD, a large database of anonymised electronic health records from UK primary care, contains patient-level data covering approximately 7% of the UK population (14). CPRD data have been validated extensively and are representative of the UK population in terms of age, sex (14), and ethnic background (15). We included patients of any age if their records were acceptable for research purposes (a data quality indicator provided by CPRD) and were registered at practices with continuous high-quality data reporting (CPRD defined up-to-standard) (16) at any time during the study period. We grouped patient data into their respective general practices.

The protocol was approved by the Independent Scientific Advisory Committee (ISAC) of the MHRA (ISAC protocol number 17_06R). Ethics approval for observational research using the CPRD with approval from ISAC was granted by a National Research Ethics Service committee (Trent MultiResearch Ethics Committee, REC reference number 05/MRE04/87).

Included tests

We examined 44 specific tests (28 laboratory, 11 imaging and five other, miscellaneous tests). The tests were chosen because they are commonly used tests or included in guidelines or in the Quality Outcomes Framework (QOF) (Supplementary file).

We grouped tests into their respective general practices, via their practice identification number. To avoid double counting, if the same code was recorded multiple times for the same patient on the same day, it was counted as only one test. Similarly, codes likely referring to the same test, or separate components of a single test (e.g. individual components of a full blood count), were grouped and counted as one test.

Statistical analysis

To identify which test was subject to the greatest between-practice variation in its use, we calculated, for each 44 tests, an unadjusted co-efficient of variation (CoV) and then an adjusted CoV.

To calculate the unadjusted co-efficient of variation, we did the following: we initially determined the number of tests ordered from each practice from April 1st 2015 to 31st March 2016. We then calculated the total person-years of observation for each general practice. Patients alive and registered for the entire year contributed 1 person-year of observation to the total. Patients who were born, died, registered, or deregistered during the year were included, but their contribution to the person-year calculation was adjusted proportionately (e.g. a patient who was registered and alive for only 6 months contributed 0.5 person-years).

We then calculated the mean unadjusted rate of use for each specific test across all 444 general practices, we also calculated the corresponding standard deviation. We used these two numbers to calculate the unadjusted CoV (standard deviation/mean x 100) (17). The use of CoV facilitates a direct comparison of the variation in use between tests controlling for differences in sample size. It is expressed as a percentage (the ratio of the standard deviation to the mean), with larger percentages reflecting greater variation.

To calculate the adjusted CoV, we constructed a generalised linear model with Poisson errors to estimate the number of tests ordered from each general practice adjusted for practice differences in patient age, sex and deprivation. We constructed 44 Poisson models for each test. The age covariate represents the median age of each general practice, the sex covariate represents the proportion of female patients in each practice and the deprivation covariate represents the practice-level Index of Multiple Deprivation (IMD) deciles. We constructed Poisson models to adjust for differences in patient demographics (age, sex and IMD) between general practices. We did not construct Poisson models to compare the predictive ability of patient demographics on the rates of test use.

We then calculated the adjusted rate of use for each test in every general practice by dividing the adjusted number of tests by the person-years for each general practice (the same process we followed to calculate the unadjusted rate of use). The adjusted rates were used to calculate the adjusted CoV for each test, as described above (17). We ranked tests according to their CoV. We present both the unadjusted and adjusted CoV in the supplementary file, but only the adjusted CoV in the main manuscript.

To identify the tests that had both a high rate of use and a high CoV we calculated the overall median rate of test use and the overall median adjusted CoV. We then classified tests into four categories: 1. High variability, low rate, 2. High variability, high rate, 3. Low variability, low rate, or 4. Low Variability, High rate. These categories reflect a test's measure in relation to the median value, e.g. high variability, low rate refers to tests with a co-efficient of variation above the median co-efficient of variation, but a rate of test use below the median rate of test use.

Role of the funding source

This study was funded by an independent grant from the National Institute for Health Research (NIHR) School of Primary Care Research (Grant reference number: 386). Independent expert peer reviewers provided feedback on the grant application underpinning this study but had no further role in study design, data collection, analysis, interpretation, or drafting of the manuscript.

Results

Data from 444 general practices contributed to the CPRD from April 1st 2015 to March 31st 2016. In total 16,496,218 tests were ordered for 4,078,091 patients over 3,311,050 person-years. The median age of patients was 40 (IQR: 21 to 58) and the median percentage of females was 50.6% (IQR: 49.7% to 51.3%). The total number of patients that each practice contributed varied (median: 6,955 (IQR: 4,374 to 9,905) and deprivation differed between included general practice, ranging from 1 to 10 (median: 6 (IQR: 4 to 8)).

Tests with the most variation in use

Figure 1 shows the rank order of the most to least variable tests. The adjusted CoV varied from 158% (95%CI: 153% to 163%) for non-illicit drug monitoring tests (urine, blood or serum, for instance serum digoxin or lithium) to 5.6% (95%CI: 5.4% to 5.8%) for testosterone tests. Urine Microalbumin (52% (95%CI: 49.9% to 53.2%)), Pelvic CT (51% (95%CI: 50% to 53%)) and Pap smear (49% (95%CI: 48% to 51%) were the second, third and fourth most variable tests. Drug monitoring, pelvic CT and pap smear tests were also the most variable laboratory, imaging and miscellaneous tests respectively. The median co-efficient of variation was 22.7% (IQR: 14.8% to 31.0%). These measures represent the between-practice variation in test ordering adjusted for age, sex and deprivation differences between practices (Appendix Table 1).

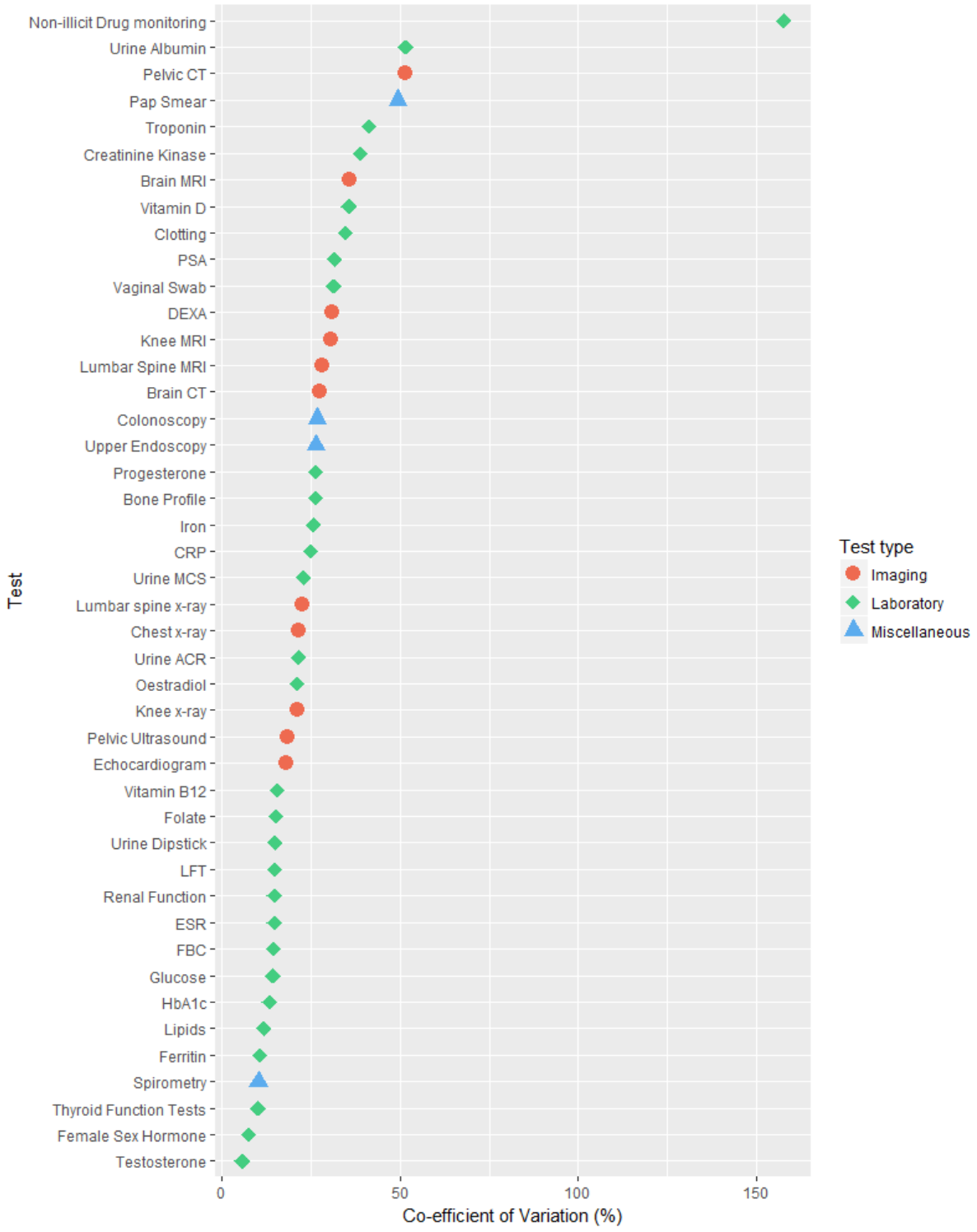

Figure 1 Rank order of variability of tests, adjusted for age, sex and deprivation. CT = Computer Tomography, MRI = Magnetic Resonance Imaging, PSA = Prostate Specific Antigen, DEXA = Dual-energy X-ray absorptiometry, CRP = C-reactive Protein, MCS = Microscopy, culture and sensitivities, ACR = Albumin-creatinine ratio, ESR = Erythrocyte sedimentation rate, LFT = Liver Function Tests, FBC = Full Blood Count.

Appendix Figure 1 shows the adjusted and unadjusted coefficients of variation for the 44 specific tests and Appendix Table 1 presents the difference between adjusted and unadjusted coefficients of variation for each specific test. Figure 2 shows an example of the adjusted and

unadjusted rate of test use (CRP), measured against the respective person-years. This figure shows how the rates of CRP use for each 444 general practices is adjusted in accordance to their practice demographic (age, sex and deprivation). Similar graphs for the other 43 tests are displayed in appendix.

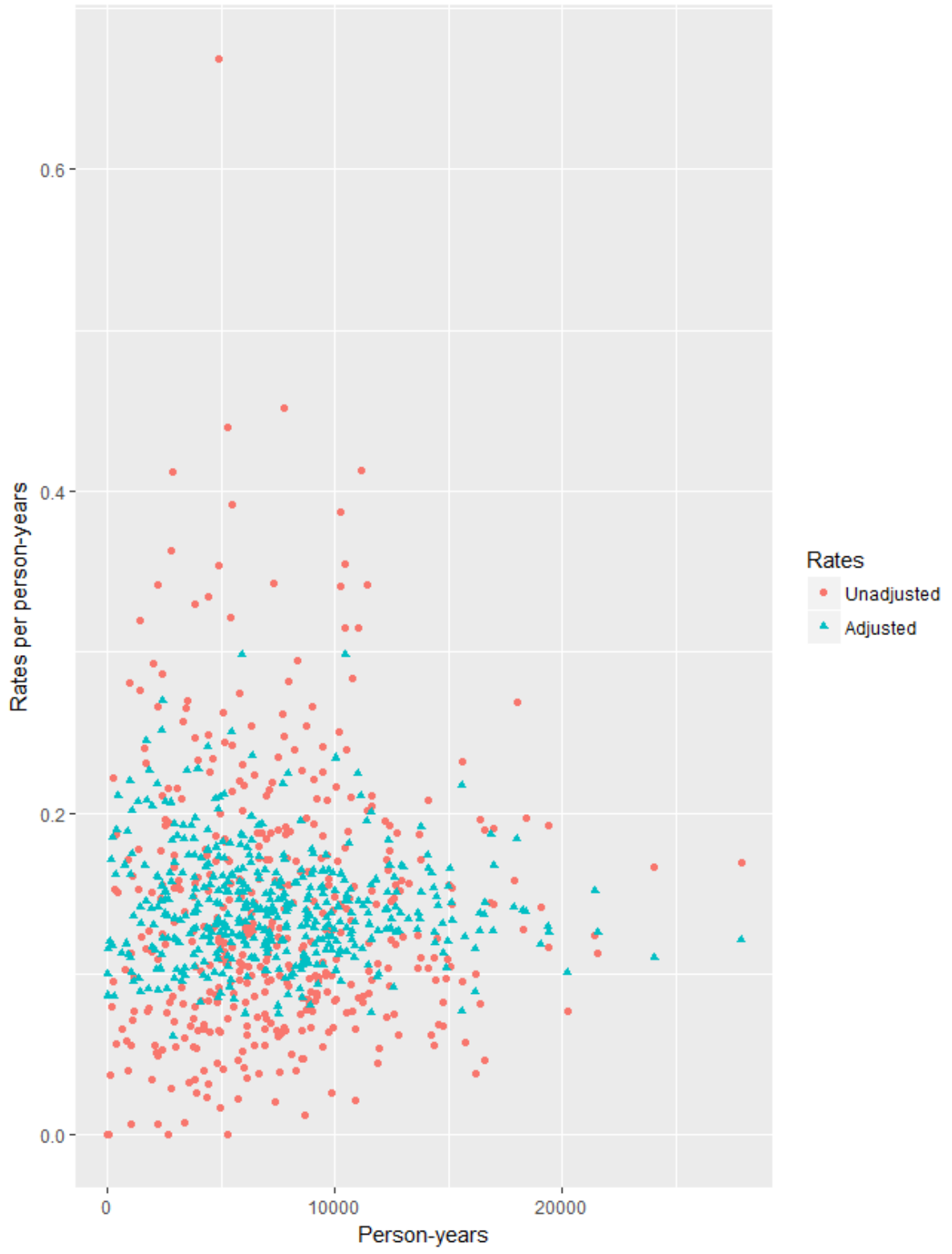

Figure 2 Adjusted and unadjusted rates for C-reactive Protein (CRP) use. All 444 practices are represented by one red and one blue data point.

Tests with a high rate of use and high variability of use

Figure 3 plots the adjusted co-efficient of variability of each test against its respective rate. The median rate was 167.3 (per 10,000 person-years) and the median CoV was 22.7. Most tests were classified as high variability, low rate (n = 15, 34%) or low variability, high rate (n = 15, 34%). Seven tests were classified as high variability and high rate (Clotting, Vitamin D, Urine Albumin, Prostate Specific Antigen (PSA), Bone profile, Urine MCS and C-reactive Protein (CRP)). The remaining seven tests were classed as low variability and low rate.

Four miscellaneous tests were in the same category (high variability, low rate: colonoscopy, upper endoscopy, pap smear, and vaginal swab), with only one (spirometry) classed as low variability, high rate. Ten of the eleven imaging tests had rates of use below the median; four of these tests were classified as low variability and low rate (Pelvic Ultrasound, lumbar x-ray, knee x-ray and Echocardiogram), while the other six were classified as high variability, low rate (pelvic CT, knee MRI, DEXA, brain MRI, brain CT and lumbar spine MRI). Only one imaging test had an adjusted rate above the median (chest x-ray), which was classified as low variability, high rate. Appendix Table 2 shows the classification of all tests.

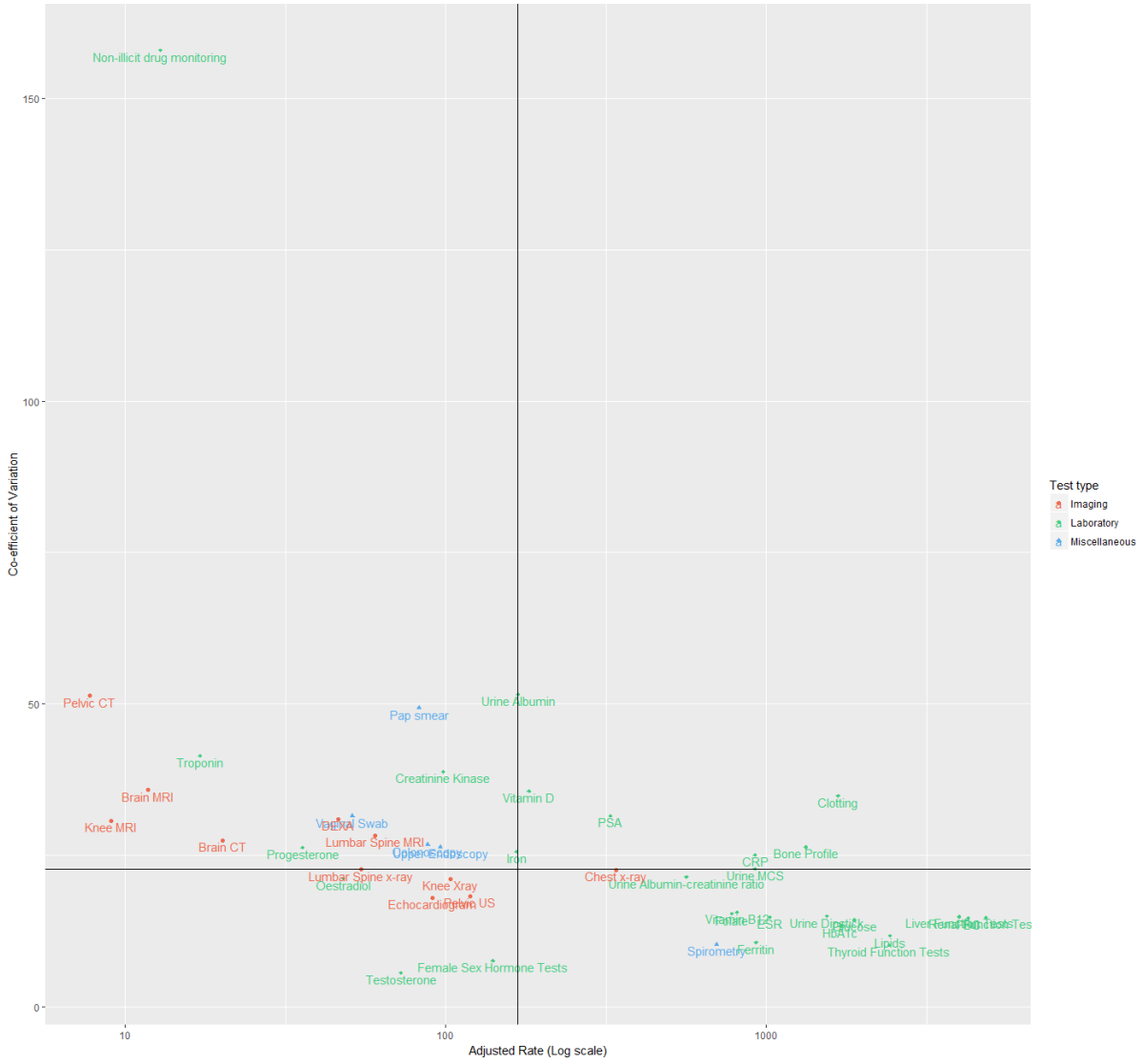

Figure 3 Variability and rates of tests. The vertical line represents the median rate of test use and the horizontal line represents the median co-efficient of variation. Median rate (vertical line) = 167.3. Median CoV (horizontal line) = 22.7

Discussion

We present a ranking of 44 primary care tests based on the between-practice variation in their use. We analysed over 16 million tests from 444 general practices and ranked tests by their adjusted coefficient of variation. The test subject to the greatest variation was non-illicit drug monitoring tests (urine, blood or serum), urine microalbumin, Pelvic CT, and Pap smear. We also identified seven tests with both a rate of ordering and a coefficient of variation above average: Clotting, Vitamin D, Urine Albumin, PSA, Bone profile, Urine MCS and CRP.

Strengths and limitations in relation to previous research

Our analysis adjusted for demographic differences between practices, however there may be valid reasons to explain the residual variation we present. Previous work has suggested that differences in disease prevalence, patient choice, data artefact (differences in data quality), resource availability, local policy and guidelines, and service configurations may also contribute to variation in healthcare resource use (10). Other previous research suggests further reasons, not all justifiable. The influence of local key-opinion leaders (18) - such as a hospital consultant preferring a one test over another - and the variation in management of uncertainty among general practitioners (19) have both been suggested as contributors.

We used the conventional statistical analysis for count data; a Poisson regression model. However, we used the outputs of this model in a less conventional manner. The aim of this paper was to determine which tests were subject to the most between-variation practice in their use once patient demographics between practices had been accounted for. We did not use the Poisson models to determine and compare the predictive ability of our covariates (patient demographics). As is expected when analysing health care data (20), the model accounted for some, but not all, of the variation in test use. This residual variation – "overdispersion" – represents the variation in test use once patient demographics between practices had been accounted for. We ranked tests by their residual variation; variation in use that persisted despite adjustment of patient demographic differences between practices.

A strength of our study is our examination of all types of tests (imaging, laboratory and miscellaneous), inclusion of many tests and our use of the appropriate statistical methods to quantify variation. One previous study has presented a ranking of primary care tests on their between-practice variation (13). This study only examined laboratory tests and included a smaller number of tests (29) in a smaller sample of patients. This study also ranked test variation by their standard deviation (SD). We preferred CoV to SD because SD is affected by the rate of testing (sample size). We found that tests that are most commonly ordered are more likely to have higher standard deviations. The use of SD to rank the between-practice variation may make tests that are ordered more commonly appear to have higher between-practice variation. A limitation of our use of CoV is that it may overestimate variation in low ordering tests; to try and mitigate this we presented both the tests with the greatest between-practice variation and the tests with a rate of ordering and a CoV above average. We believe tests with both a high CoV and high rate of ordering should be the focus of future academic and policy work.

A further strength of our study is the use of high-quality, validated electronic health record data and the identification of tests that are subject to the greatest between-practice variation. Most previous research exploring geographical variation in healthcare resource use has focused on identifying regions that order a greater or lesser number of tests or treatments compared to the national average (11,21–24).

Implications for practice and policy

The wide between-practice variation in test use we present is unlikely to be explained entirely by clinical indication. We present a list of common and important primary care tests ranked by their between-practice variation. Policy makers must decide if the residual variation we present is warranted, and if it is not, understand why this variation exists and what can be done to mitigate it. Between-practice variation and, more broadly, geographical variation have long been used to highlight potential over or underuse of healthcare resources (23,25–27). Our ranking of tests can direct policy makers to the primary care tests most likely subject to overuse – the use of a test when it will not result in patient benefit - or underuse – the failure to use a test when it would result in patient benefit. However, it should be noted the variation we present does not directly consider individual patient data, nor the clinical indications for test use. As such, our results can be considered a potential, not definitive, indicator of over and underuse.

In some cases, there are content-specific reasons to explain the between-practice variation in test use. For instance, the notable between-practice variation in the use of clotting and drug monitoring tests may reflect regional differences in drug use. In UK primary care, there has been an increase in the use of novel oral anticoagulants (NOAC) (also known as direct acting oral anticoagulants (DOACs) (28); from 2009 to 2015, there was a 17-fold increase in NOAC use (28). However, there is marked geographical variation in their use (29). This variation may reflect the non-specific NICE guidance; it states that patients with atrial fibrillation can be anti-coagulated with "apixaban, dabigatran, rivaroxaban or a vitamin K antagonist" (30). However, this guidance is now out-of-date compared to more recent evidence. A 2017 systematic review and network meta-analysis concluded that "the risk of all-cause mortality was lower with all DOACs" and "several DOACs are of net benefit compared with warfarin" (31). With clear guidance, reflecting the underlying evidence, it is plausible that geographical variation in clotting tests would diminish.

Similarly, the variability of drug monitoring tests is likely to be related to regional differences in disease prevalence. Drug monitoring tests include tests for tacrolimus, cyclosporin, salicylate, lamotrigine, lithium and gentamicin (among others). All of these tests, individually, had low rates of use. Lastly, it should be noted that tests can be directly wasteful, but can also contribute to healthcare costs indirectly, for instance via incidental imaging findings (32).

Future research

It would be advantageous for future studies to investigate variation using a different unit of an analysis. We chose to investigate variation at a practice level, however future studies could investigate variation at a patient-level or at a regional level. We chose to investigate at a practice level as previous literature suggest practice level factors contribute substantial to healthcare variation (10,18). Differences in disease prevalence, cultural attitude to tests and risks, local key-opinion leaders, resource availability, local policy and guidelines, and service configurations have all been suggested as practice-level contributors to variation.

A similar analysis aggregated at a regional, rather than practice, level may provide further insight into unwarranted variation. It is plausible that our analysis at a practice level may be too sensitive to variation in disease prevalence, this may in part explain non-illicit drug testing as an outlier. However, the aggregation of data at a regional level may obfuscate true, unwarranted variation. Furthermore, the CPRD only allows practices to be identified at a broad regional level (e.g. within Wales). Conversely, future research that analyses data at an individual patient level may provide more nuanced insight into variation, but risks being overly sensitive; making the distinction between warranted and unwarranted variation more

difficult. Nevertheless, we would welcome any further studies using the aforementioned analyses.

Furthermore, beyond adjustment for demographic differences, we could not directly determine the appropriateness of the between-practice variation we noted. Future research studies should aim to determine if the tests with the greatest between-practice variation are also subject to the greatest underuse and overuse. This research should ideally involve individual patient data (IPD) either in the form of notes review, or IPD data audit (33), commonly against guidelines (34). Some of our team are involved in delivering OpenPathology.net (23); an open data tool (like OpenPrescribing.net) that provides easy access to various analytic approaches identifying test- ordering behaviour in primary care. This tool will continue our work exploring temporal trends on a live interface.

Conclusions

There is wide variation among commonly used tests, which is unlikely to be explained by clinical indication, and since £3 billion annually are spent on tests this represents considerable resource use variation and inefficient management for the NHS.

Declarations

Ethical approval

The protocol was approved by the Independent Scientific Advisory Committee (ISAC) of the MHRA (ISAC protocol number 17_06R). Ethics approval for observational research using the CPRD with approval from ISAC was granted by a National Research Ethics Service committee (Trent MultiResearch Ethics Committee, REC reference number 05/MRE04/87).

Consent for publication

Not applicable

Availability of data and materials

The data that support the findings of this study are available from the Clinical Practice Research Datalink but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Clinical Practice Research Datalink.

Competing interests

The authors declare that they have no competing interests.

Funding

National Institute for Health Research School of Primary Care Research (Award Number 386)

Author's contributions

JOS, CH, RP, BG and FDRH conceived the idea for the research. JOS, RP, CB and SS designed the study, which was further refined by GP experts CH, FDRH, PL and CS. JOS drafted the protocol, which all authors contributed and revised critically. JOS and SS were responsible for data management and JOS, SS and RP did the statistical analyses. JOS drafted

the manuscript, to which all authors contributed, revised critically and approved. JOS is the guarantee.

Acknowledgements

This study was funded by an independent grant from the National Institute for Health Research (NIHR) School of Primary Care Research (Grant reference number: 386). JOS is a doctoral student supported by the Clarendon Fund. FDRH is a general practitioner and research lead with the Modality Partnership and Director of the National Institute for Health Research (NIHR) School for Primary Care Research. FDRH acknowledges part funding support from the NIHR School for Primary Care Research, the NIHR Oxford BRC, and the NIHR CLAHRC Oxford. CS is a member of the NIHR Health Services and Delivery Research Board and acknowledges support from NIHR CLAHRC West and NHS Bristol Clinical Commissioning Group. CH has received expenses and fees for his media work including BBC Inside Health. He holds grant funding from the NIHR, the NIHR School of Primary Care Research, The Wellcome Trust, and the WHO. He has also received income from the publication of a series of toolkit books published by Blackwells. With some international partners, CEBM jointly runs the EvidenceLive Conference and the Overdiagnosis Conference, which are based on a non-profit model. CB is partially supported by the NIHR Biomedical Research Centre, Oxford.

References

1. Kleinert S, Horton R. From universal health coverage to right care for health. Lancet [Internet]. Elsevier Ltd; 2017;390(10090):101–2. Available from: http://dx.doi.org/10.1016/S0140-6736(16)32588-0

2. Fisher ES, Bynum JP, Skinner JS. Slowing the growth of health care costs--lessons from regional variation. N Engl J Med [Internet]. 2009 Feb 26 [cited 2016 May 3];360(9):849–52. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2722744&tool=pmcentrez&rendertype=abstract

3. Alderwick H, Robertson R, Appleby J, Dunn P, Maguire D. Better value in the NHS The role of changes in clinical practice. 2015;(July).

4. The King's Fund. The NHS budget and how it has changed [Internet]. London; 2017. Available from: https://www.kingsfund.org.uk/projects/nhs-in-a-nutshell/nhs-budget

5. Martin AB, Hartman M, Washington B, Catlin A. National Health Spending: Faster Growth In 2015 As Coverage Expands And Utilization Increases. Health Aff [Internet]. Health Affairs; 2017 Jan 1;36(1):166–76. Available from: https://www.healthaffairs.org/doi/abs/10.1377/hlthaff.2016.1330

6. Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, Heath I, et al. Evidence for overuse of medical services around the world. Lancet [Internet]. Elsevier Ltd; 2017;6736(16):1–13. Available from: http://dx.doi.org/10.1016/S0140-6736(16)32585-5%5Cnhttp://linkinghub.elsevier.com/retrieve/pii/S0140673616325855

7. Hobbs FDR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. Lancet [Internet]. Elsevier; 2016 Jun [cited 2016 Aug 19];387(10035):2323–30. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0140673616006206

8. Centers for Disease Control and Prevention, National Center for Health Statistics. National Ambulatory Medical Care Survey: 2012 Summary Tables. 2012;5. Available from: http://www.cdc.gov/nchs/data/ahcd/namcs_summary/2010_namcs_web_tables.pdf

9. Wennberg JE. Tracking Medicine. New York: Oxford University Press; 2010.

10. Appleby J, Raleigh V, Frosini F, Bevan G, Gao H, Lyscom T. Variations in health care: The good, the bad and the inexplicable. 2011.

11. NHS Right Care. Diagnostics: The NHS Atlas of Variation in Diagnostic Services. 2012;(November):220.

12. Lord Carter of Coles. Report of the review of NHS pathology services in England [Internet]. 2006 [cited 2018 Jan 5]. Available from: http://webarchive.nationalarchives.gov.uk/20130107105354/http:/www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_091984.pdf

13. Busby J, Schroeder K, Woltersdorf W, Sterne JAC, Ben-Shlomo Y, Hay A, et al. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices: Using routine data to identify research priorities. Br J Gen Pract. 2013;63(609):256–66.

14. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol [Internet]. Oxford University Press; 2015 Jun [cited 2016 Aug 21];44(3):827–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26050254

15. Mathur R, Bhaskaran K, Chaturvedi N, Leon DA, vanStaa T, Grundy E, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. J Public Health (Oxf). 2014;36(4):684–92.

16. Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. Ther Adv drug Saf [Internet]. SAGE Publications; 2012 Apr [cited 2016 Aug 21];3(2):89–99. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25083228

17. Armitage P, Berry G. Statistical Methods in Medical Research. 1994. 40 p.

18. Wennberg JE. Time to tackle unwarranted variations in practice. BMJ. 2011;342:d1513.

19. Morgan M, Jenkins L, Ridsdale L. Patient pressure for referral for headache: A qualitative study of GP's referral behaviour. Br J Gen Pract. 2007;57(534):29–35.

20. Spiegelhalter DJ. Handling over-dispersion of performance indicators. Qual Saf Heal Care [Internet]. 2005;14(5):347–51. Available from: http://qualitysafety.bmj.com/lookup/doi/10.1136/qshc.2005.013755

21. NHS Rightcare. The NHS Atlas of Variation in Healthcare [Internet]. 2015. Available from: http://www.rightcare.nhs.uk/atlas/downloads/2909/RC_nhsAtlasFULL_LOW_290915.pdf

22. Care TAC on S and Q in H. Australian Atlas of Healthcare Variation. 2013.

23. O'Sullivan JW, Heneghan C, Perera R, Oke J, Aronson JK, Shine B, et al. Variation in diagnostic test requests and outcomes: a preliminary metric for OpenPathology.net. Sci Rep [Internet]. 2018;8(1):4752. Available from: https://doi.org/10.1038/s41598-018-23263-z

24. Wennberg JE. UNDERSTANDING GEOGRAPHIC VARIATIONS IN HEALTH CARE DELIVERY. N Engl J Med. 1999;

25. Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, Heath I, et al. Evidence for overuse of medical services around the world. Lancet [Internet]. Elsevier Ltd; 2017;6736(16):1–13. Available from: http://dx.doi.org/10.1016/S0140-6736(16)32585-5%5Cnhttp://linkinghub.elsevier.com/retrieve/pii/S0140673616325855

26. Glasziou P, Straus SE, Brownlee S, Trevena L, Dans L, Guyatt G, et al. Evidence for underuse of effective medical services around the world. Lancet [Internet]. 2017;390:169–77. Available from: http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(16)30946-1.pdf

27. Chalmers K, Pearson S, Elshaug AG. Quantifying low-value care: a patient- versus service-centric lens. BMJ Qual Saf. 2017;26:855–8.

28. Loo SY, Dell'Aniello S, Huiart L, Renoux C. Trends in the prescription of novel oral anticoagulants in UK primary care. Br J Clin Pharmacol. 2017;83(9):2096–106.

29. Millett D. NOAC prescribing varies 16-fold between CCG areas [Internet]. GP Online. 2016. Available from: https://www.gponline.com/noac-prescribing-varies-16-fold-ccg-areas/cv-thromboembolic-disorders/atrial-fibrillation/article/1394082

30. NICE. NICE guideline: Atrial Fibrillation : management [Internet]. 2014. Available from: https://www.nice.org.uk/guidance/cg180/resources/atrial-fibrillation-management-pdf-35109805981381

31. López-López JA, Sterne JAC, Thom HHZ, Higgins JPT, Hingorani AD, Okoli GN, et al. Oral anticoagulants for prevention of stroke in atrial fibrillation: systematic review, network meta-analysis, and cost effectiveness analysis. Bmj [Internet]. 2017;j5058. Available from: http://www.bmj.com/lookup/doi/10.1136/bmj.j5058

32. O'Sullivan JW, Muntinga T, Grigg S, Ioannidis JPA. Prevalence and outcomes of incidental imaging findings: Umbrella review. BMJ. 2018;361(k2387).

33. O'Sullivan JW, Albasri A, Nicholson B, Perera R, Aronson J, Roberts N, et al. Overtesting and undertesting in primary care: a systematic review and meta-analysis. BMJ Open. 2018;8:e018557.

34. O'Sullivan JW, Albasri A, Koshiaris C, Aronson JK, Heneghan C, Perera R. Diagnostic test guidelines based on high-quality evidence had greater rates of adherence : a meta-epidemiological study. J Clin Epidemiol. Elsevier Inc; 2018;103:40–50.

Manuscript with tracked changes

Supplementary Material

Supplementary Material

Response to comments