Smoking, blood cells, and myeloproliferative neoplasms: meta-analysis and mendelian randomization of 2.3 million people

Short title: Smoking, blood cells, and myeloproliferative neoplasms

Nimesh Amrith Jayasuriya[1,2], Alisa D. Kjaergaard[1,3], Kasper M. Pedersen[4,5,6], Anders Lindholm Sørensen[4,5], Marie Bak[5], Morten Kranker Larsen[5,7], Børge G. Nordestgaard[4,8], Stig E. Bojesen[4,8], Yunus Çolak[4,8], Vibe Skov[5], Lasse Kjær[5], Janne Schurmann Tolstrup[9,10], Hans Carl Hasselbalch[4,5], Christina Ellervik[1,4,11,12]

[1]Department of Laboratory Medicine, Boston Children's Hospital, Boston, USA
[2] School of Medicine, University of Glasgow, Glasgow, United Kingdom
[3]Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark
[4]Faculty of Health and Medical Sciences, University of Copenhagen, Denmark
[5]Department of Hematology, Zealand University Hospital, Roskilde, Denmark
[6]Department of Clinical Biochemistry, Copenhagen University Hospital, Herlev and Gentofte Hospital, Herlev, Denmark
[7]Department of Science and Environment, Roskilde University, Denmark
[8]Department of Clinical Biochemistry and the Copenhagen General Population Study, Copenhagen University Hospital, Herlev and Gentofte Hospital, Herlev, Denmark
[9]Faculty of Health and Medical Sciences, University of Southern Denmark, Denmark
[10]National Institute of Public Health, University of Southern Denmark
[11]Department of Production, Research, and Innovation; Region Zealand, Sorø, Denmark
[12]Division of Pathology, Faculty of Medicine, Harvard Medical School, Boston, USA

Word count: main text (3001 of 3000 words). Abstract (200 of 200 words).
6 figures, 1 table, 37 supplementary figures, 14 supplementary tables.
This study has been partly presented as a poster at 59th Annual American Society of Hematology conference in December 2017.

Correspondence: Christina Ellervik, MD, PhD, DMSci. Affiliated Associate Professor, University of Copenhagen, Denmark. Assistant Professor, Harvard Medical School, Boston, USA. Associated Medical Director, Department of Laboratory Medicine, 300 Longwood Avenue, Boston Children's Hospital, 02115 Boston, USA. Cell-phone: +1 520 955 4696. email: christina@ellervik.dk, christina.ellervik@childrens.harvard.edu, Fax: 617-730-0383

Abstract

Meta-analyses and mendelian randomization(MR) may clarify the associations of smoking, blood cells, and myeloproliferative neoplasms(MPN). We investigated the association of smoking with blood cells in the Danish General Suburban Population Study(GESUS,N=11,083), meta-analyses(including GESUS) of 92 studies(N=531,741), and MR of smoking variant *CHRNA3*(rs1051730[A]) in UK Biobank, and with MPN in a meta-analysis of 6 studies(N(total/cases):1,425,529/2187), totaling 2,307,745 participants.

In the meta-analysis the random effects SMD in current smokers vs. non-smokers was 0.82(0.75-0.89,p=$2.0*10^{-108}$) for leukocytes, 0.09(-0.02-0.21,p=0.12) for erythrocytes, 0.53(0.42-0.64,p=$8.0*10^{-22}$) for haematocrit, 0.42(0.34-0.51,p=$7.1*10^{-21}$) for haemoglobin, 0.19(0.08-0.31,p=$1.2*10^{-3}$) for MCH, 0.29(0.19-0.39,p=$1.6*10^{-8}$) for MCV, and 0.04(-0.04-0.13,p=0.34) for platelets with trends for ever/ex-/current smokers, light/heavy smokers, and female/male smokers. Analyses presented high heterogeneity but low publication bias. Per allele in *CHRNA3,* cigarettes per day in current smokers were associated with increased blood cell counts (leukocytes, neutrophils), MCH, RDW, and MCV. Pooled fixed effects OR for MPN was 1.44(95%CI:1.33-1.56;p=$1.8*10^{-19}$;I^2=0%) in current smokers, 1.29(1.15-1.44;p=$8.0*10^{-6}$;I^2=0%) in ex-smokers, 1.49(1.26-1.77;p=$4.4*10^{-6}$;I-square=0%) in light-smokers, and 2.04(1.74-2.39,p=$2.3*10^{-18}$;I^2=51%) in heavy-smokers compared with non-smokers. Smoking is observationally and genetically associated with blood cell counts in current and ex-smokers and observationally with risk of MPN.

Keywords:

Blood cells, tobacco smoking, myeloproliferative neoplasms, essential thrombocytaemia, polycytaemia vera, myelofibrosis, meta-analysis, haematocrit, haemoglobin, leukocytes, erythrocytes, thrombocytes

Abbreviations (in alphabetic order):

CBC: Complete blood count

ET: Essential thrombocytaemia

GESUS: The Danish General Suburban Population Study

HCT: Haematocrit

Hgb: Haemoglobin

IRF: Immature reticulocyte fraction

MCH: Mean corpuscular haemoglobin

MCHC: Mean corpuscular haemoglobin concentration

MCV: Mean corpuscular volume

MF: Myelofibrosis

MPN: Myeloproliferative neoplasms

MPV: Mean platelet volume

MRV: Mean reticulocyte volume

NLR: Neutrophil lymphocyte ratio

PCT: Platelet crit

PDW: Platelet distribution width

PLT: Platelet

PV: Polycytaemia vera

RBC: Red blood cell

RDW: Red blood cell distribution width (CV: Coefficient of variation (%))

WBC: White blood cell

Introduction

The classic Philadelphia-negative chronic myeloproliferative neoplasms (MPN) are acquired stem cell diseases that include essential thrombocytaemia (ET), polycytaemia vera (PV), primary myelofibrosis (MF) and unclassifiable MPN (MPN-U)(Spivak 2017). Patients with MPN have an increased comorbidity burden and exhibit shorter survival compared to the general population(Hultcrantz, *et al* 2015). Chronic inflammation may be a driving force for clonal evolution and disease progression from the early cancer stage with elevated cell counts (ET, early prefibrotic myelofibrosis, PV) to the advanced burnt-out phase with severe myelofibrosis (Campbell, *et al* 2005, Frederiksen, *et al* 2011, Hasselbalch 2012, Hasselbalch and Bjorn 2015, Larsen, *et al* 2007).

Smoking associated toxic substances in tobacco are highly potent inflammatory stimuli and associated with tumorigenesis and clonal hematopoiesis (Hasselbalch 2015). Smoking associates with increased blood cells(Billimoria, *et al* 1975, Dotevall, *et al* 1987, Eisenga, *et al* 2018, Lao, *et al* 2009, Lowe, *et al* 1992, Parry, *et al* 1997, Smith, *et al* 2003, Wannamethee, *et al* 2005) (full list of references in Supplementary Material) and MPN (Kroll, *et al* 2012, Leal, *et al* 2014, Lindholm Sorensen and Hasselbalch 2016, Pasqualetti, *et al* 1997, Pedersen, *et al* 2018) in some but not all observational studies, likely due to differences in study size, effect size and direction, year published, sex and age distribution, geographical region, smoking status (current vs. ex-smoker), smoking duration, smoking intensity (heavy vs. light), and ethnicity. Smoking and MPNs share several clinical and biochemical characteristics including a high risk of thrombosis, partly explained by an elevated haematocrit and leukocytosis and *in vivo* leukocyte, platelet and endothelial activation (Hasselbalch 2015). Yet, no previous meta-analyses of smoking, differences in complete blood counts (CBC), and MPN have been published.

As observational studies may be prone to reverse causation, the genome-wide association (GWAS) SNP *CHRNA3* (The Cholinergic Receptor Nicotinic Alpha 3 Subunit, rs1051730 [A]), which associates with number of cigarettes smoked per day(Tobacco and Genetics 2010) (Pedersen, *et al* 2019) may provide an opportunity to assess associations to CBC through an instrumental variable approach termed Mendelian Randomization (MR).

We investigated the association of cigarette smoking with CBC in a Danish observational study (N=11,083), meta-analyses of 92 studies (N=531,741), and MR analysis of *CHRNA3* in UKbiobank (N=350,475), and the association of smoking with MPN in a meta-analysis of 6 studies (N=1,425,529), totaling 2,307,745 participants.

Methods

The Danish General Suburban Population Study (GESUS) is a cross-sectional general population study of 21,205 adults (participation rate 43%) from 2010-2013(Bergholdt, *et al* 2013, Cordua, *et al* 2019). We included 11,083 healthy participants (mean age was 52 years (SD: 12), 53.5% women) with no cardiovascular disease or cancer. CBC were measured on fresh EDTA whole blood on Sysmex XE-5000 within 24h of blood draw. We calculated adjusted means for CBC by smoking status (never, current, ex-, and ever smoker) from multivariable linear regression models adjusted for potential confounders that a priori could be associated with tobacco smoking and/or blood cell values (Supplementary methods).

The meta-analysis protocol was registered at the International Prospective Register of Systematic Reviews (Prospero), https://www.crd.york.ac.uk/PROSPERO/, CRD42018102571(Page, *et al* 2018). A systemic literature search with assistance from

Countway Library at Harvard Medical School was conducted on the databases Pubmed, EMBASE, Web of Science and World Health Organisation Global Health Library for all relevant papers published on smoking, complete blood cell count indices, and myeloproliferative neoplasms until June 11th, 2018 (**Supplementary methods)**. An article was relevant if it originated from cross-sectional, case control, retrospective, population, prospective, or trial studies at baseline and reported original data on CBC or MPN. For CBC, we calculated the pooled effect size using the unitless standardized mean difference (SMD, 95% CI) and the pooled non-standardized (weighted) mean differences (WMD, 95% CI). For CBC, we used the random-effects model (DerSimonian and Laird) due to heterogeneity. For MPN, we calculated the fixed effects pooled odds ratio (95% CI) as there was no study heterogeneity. Heterogeneity, sensitivity analyses, and publication bias are described in **Supplementary Methods**. We assessed double-sided p-values with Bonferroni correction for multiple comparisons by dividing the p-values with 24 cell type measurements to obtain an overall global significance of double-sided $p<0.002$.

We used instrumental variable analysis in a *two-sample Mendelian randomization* approach with summary-level exposure of cigarettes per day using Cholinergic Receptor Nicotinic Alpha 3 Subunit *CHRNA3* (rs1051730 [A])(Liu, *et al* 2010). We extracted genetic summary statistics for CBC from UK Biobank (Bycroft, *et al* 2018). CBC were measured on fresh EDTA whole blood on Beckman Coulter LH750 within 24h of blood draw.

CBC was reported using SI units(Brereton, *et al* 2016). MPN was defined as any of the following: ET, PV, pre-MF, MF, and MPN-U, by ICD coding from registries, WHO coding by haematologists, or by bone marrow examination by pathologists(Barbui, *et al* 2018).

Results

Complete Blood Count

The Danish General Suburban Population Study (GESUS)

Characteristics of the study participants based on smoking status are shown in **Supplementary Table 1**. In current smokers, ex-smokers, or ever smokers compared to never smokers, leukocyte counts were increased (**Figure 1, Supplementary Figures 1-2**); results were similar for all differential subpopulations, except for NLR and basophils. In current smokers, ex-smokers, or ever smokers compared to never smokers, erythrocyte counts were decreased (**Figure 1, Supplementary Figures 1-2**). In current smokers compared to never smokers, haematocrit, haemoglobin, MCH, MCV, and RDW were increased (**Supplementary Figure 1**), but in ex-smokers, haematocrit and haemoglobin were decreased (**Supplementary Figure 2**). In current smokers, ex-smokers, or ever smokers compared to never smokers, platelet counts and PCT were increased (**Figure 1, Supplementary Figures 1-2**), but MPV and PDW were not different.

Meta-analysis

For the meta-analysis, we included 92 study populations (including GESUS) on CBC published from 1974-2018 comprising 531,741 individuals with 214,703 current smokers, 90,174 ex-smokers, and 222,600 never or non-smokers between the ages of 15 and 99 years from 33 countries (**Figure 2, Table 1, Supplementary Table 2**). Males were overrepresented in the meta-analyses with a male:female ratio of 6:1.

Random effects non-standardized weighted mean differences (WMD) in current smokers vs. non-smokers per cell type are shown in **Supplementary Figures 3-26**. The random effects SMD of leukocytes in current smokers was 0.82(0.75-0.89, $p=2.0*10^{-108}$) overall (**Figure 3**) equivalent to a mean difference (WMD) of $1.25*10^{9}$/L (**Supplementary Figure 3**). Compared with non-smokers, the SMD of leukocytes was 1.56(1.34-1.79, $p=7.2*10^{-41}$) in current heavy smokers, 0.87(0.69-1.05,$p=2.2*10^{-21}$) in current light smokers, 0.84(0.75-0.93,$p=1.1*10^{-72}$) in current male smokers, and 0.65(0.57-0.73,$p=3.5*10^{-60}$) in current female smokers (**Supplementary Figures 27-30**). Compared with never smokers, the SMD was 0.16(0.11-0.21, $p=8.0*10^{-12}$) in ex-smokers, and 0.60(0.52-0.68),$p=2.2*10^{-53}$) in ever-smokers (**Supplementary Figures 31-32**). SMD of absolute counts, but not percentages, for neutrophils, eosinophils, basophils, monocytes, and lymphocytes were similar (**Figure 3, Supplementary Figures 27-32**).

Random effects SMD of haematocrit in current smokers was 0.53(0.41-0.65, $p=6.7*10^{-19}$) overall (**Figure 3**) equivalent to a WMD of 0.02 (L/L) (**Supplementary Figure 17**). Compared with non-smokers, the SMD of haematocrit was 1.08(0.77-1.40, $p=1.4*10^{-11}$) in current heavy smokers, 0.54(0.21-0.87,$p=1.4*10^{-3}$) in current light smokers, 0.29(0.20-0.37,$p=3.2*10^{-11}$) in current male smokers, and 0.33(0.11-0.55,$p=3.3*10^{-3}$) in current female smokers. Compared with never smokers, the SMD was 0.47(0.12-0.82, $p=7.8*10^{-3}$) in ex-smokers, and 0.19(0.10-0.28),$p=7.3*10^{-5}$) in ever-smokers (**Supplementary Figures 27-32**). Results were similar and in the same direction for erythrocytes, haemoglobin, MCH, MCV, and RDW (**Figure 3, Supplementary Figures 27-32**). Results for MCHC and reticulocytes were not consistent (**Figure 3, Supplementary Figures 27-32**).

Compared with non-smokers, random effects SMD of platelets was 0.04(-0.04-0.13, p=0.34) overall in current smokers (**Figure 3**), 0.22(0.09-0.34, $p=5.2*10^{-4}$) in current heavy

smokers, 0.09(-0.05-0.23,p=0.21) in current light smokers, 0.05 (-0.06-0.16,p=0.40) in current male smokers, and 0.01 (-0.14-0.16,p=0.88) in current female smokers. Compared with never smokers, the SMD was 0.06 (0.00-0.11, p=0.04) in ex-smokers, and 0.15(0.10-0.19),p=$4.6*10^{-10}$) in ever-smokers (**Supplementary Figures 27-32**). Random effects SMD of MPV and PDW were not significant in any analyses (**Figure 3, Supplementary Figures 27-32**).

All analyses presented very high heterogeneity, reflected by the I^2 and Q-statistic. Egger's tests for publication bias revealed no publication bias except in analyses of leukocytes and platelets in current smokers. Analyses by ethnicity revealed that erythrocyte counts were higher in current smokers compared to non-smokers in Middle Eastern countries [SMD(95%CI): 0.49 (0.12,0.85), p=0.009] whereas erythrocyte counts were lower in current smokers compared to non-smokers in Western European countries [SMD(95%CI): -0.23(-0.39,-0.07), p=0.006] (p-difference=0.02) (**Supplementary Table 3**); the SMD by ethnicity was not different for leukocytes, haematocrit, haemoglobin, or platelets. Analyses by laboratory method did not reveal any differences and did not remove heterogeneity (**Supplementary Table 4**). Meta-regressions did not reveal impact of moderator variables (study size, publication year, longitude or latitude of study location, or mean ages) (**Supplementary Table 5**). Leave-one-study-out analyses for leukocytes, erythrocytes, haemoglobin, haematocrit, and platelets did not reveal any exaggerated influences of studies (**Supplementary Tables 6-10**). A sub-group analysis of all cell types in current smokers vs. never smokers revealed similar but accentuated results compared to analyses in current smokers vs. non-smokers (**Supplementary Figure 33**).

Mendelian randomization study in UK Biobank

The mean(SD) of cigarette smoking per day was 15.5(8.4) in current smokers and 18.9(10.9) in ex-smokers. The β(SE) and instrument strength F was [0.0853(0.00439), 376] per SD of cigarettes per day for *CHRNA3* (rs1051730 [A]). *CHRNA* (rs1051730[A]) associated with chronic obstructive lung disease and lung cancer in UK Biobank (**Supplementary Table 11**).

For each rs1051730[A] in *CHRNA3* in current smokers, one SD increase in cigarettes per day (~10 cigarettes per day) were associated with SD increases in leukocytes of 0.07, neutrophils of 0.09, MCH of 0.15, MCV of 0.16, RDW of 0.8, and reticulocytes of 0.07 (**Figure 3**), with results for male higher than in female, and with results persisting but attenuated in ex-smokers (**Supplementary Figures 34-36**).

Meta-analysis of Myeloproliferative Neoplasms

We included 6 studies, 1,425,529 individuals with 2187 MPN cases (0.2%) (**Figure 2, Table 1, Supplementary Table 13**) published between 1997-2018. The pooled fixed effects odds ratios for MPN were 1.44(95%CI: 1.33-1.56;p=$1.8*10^{-19}$;I^2=0%) in current smokers, 1.49(1.26-1.77;p=$4.4*10^{-6}$;I-square=0%) in light-smokers, 2.04(1.74-2.39,p=$2.3*10^{-18}$;I^2=51%) in heavy-smokers vs. non-smokers, and 1.29(1.15-1.44;p=$8.0*10^{-6}$;I^2=0%) in ex-smokers and 1.39(1.30-1.48,p=$8.3*10^{-23}$;I^2=35%) in ever smokers vs. never-smokers (**Figure 5 and 6**). There was no evidence of publication bias. Subgroup analyses for PV and ET showed similar results for current smokers, which had overlapping confidence intervals with ex-smokers and ever-smokers; there were only two studies including patients with primary MF (**Supplementary Table 14**).

Discussion

We examined the association between smoking, blood counts, and risk of MPN in 2.3 million participants. In the Danish population study and in the meta-analysis, current smoking was associated with increased leukocyte counts (including, neutrophils, eosinophils, monocytes, lymphocytes), haematocrit, haemoglobin, MCH, and MCV, and increased risk of MPN compared with non-smokers. Leukocyte counts were impacted the most relative to other cell types. Compared to non-smokers, there was a dose-trend for cell counts, indices, and MPN for ever/ex-/current smokers and light/heavy smokers, and male/female smokers. All analyses for CBC presented high between-study heterogeneity but low publication bias, whereas analyses for MPN presented low to moderate between-study heterogeneity with no publication bias. In the MR study, higher cigarette consumption in current smokers was associated with increased leucocytes, neutrophils, MCH, MCV, RDW, and reticulocytes.

The chemical compounds found in tobacco, such as nicotine, tar, carbon monoxide, and hydrogen cyanide, are associated with genotoxic, cytotoxic and cancerogenic effects(Lohler and Wollenberg 2019). Thus, the observed associations between smoking and increased CBCs could merely reflect production of cells in reaction and secondary to smoking in an otherwise healthy cell lineage. It could also reflect acquired mutations in hematopoietic stem cells in the bone marrow with clonal hematopoiesis of indeterminate potential or MPN, as the associations persisted despite smoking cessation (Baxter, *et al* 2005, Nielsen, *et al* 2013, Weinberg, *et al* 2012).

Smoking and MPNs share several clinical, biochemical and molecular characteristics, which may all be mediated by chronic inflammation and oxidative stress(Hasselbalch 2015). Smoking is associated with genotoxic effects, namely the acquired somatic driver mutation *JAK2 V617F* for MPN(Baxter, *et al* 2005, Nielsen, *et al* 2013, Weinberg, *et al* 2012), and oxidative

damage to DNA leading to increased reactive oxygen species(Prieme, *et al* 1998). The JAK-STAT and NF-κB signaling pathways, associated with carcinogenesis, are activated and perpetuated by reactive oxygen species in both smokers and in patients with MPNs(Bjorn and Hasselbalch 2015, Hasselbalch 2015). Smokers, as well as patients with MPN, have a high risk of thrombosis, which is likely explained by an elevated haematocrit and leukocytosis and *in vivo* leukocyte, platelet and endothelial activation(Hasselbalch 2015, Spivak 2017). Interestingly, patients with MPNs have increased risk of secondary cancers, particularly lung cancer and urinary tract cancers, which are also related to smoking(Frederiksen, *et al* 2011, Landtblom, *et al* 2018).

Some CBC results were discordant in the general population study, the meta-analysis, and the MR study. In both the MR study and the general population study, current smoking was inversely associated with erythrocyte counts, but in the meta-analysis current smoking associated with increased erythrocyte counts. There can be several explanations for these findings. First, smoking is associated with hypoxia as carbon monoxide binds to haemoglobin and thereby reduces the oxygen-binding capacity which theoretically would lead to increased erythrocytes as a compensatory mechanism(Goldstein 2008). Second, the lower counts of erythrocytes in current smokers compared to non-smokers observed in the general population study along with an increased heterogeneity in red cell size (RDW) could be caused by a decreased erythropoiesis, increased cell destruction, bone marrow depression, and chronic inflammation(Asgary, *et al* 2005, Li, *et al* 2017). Third, the inverse relationship between smoking and erythrocyte counts could also be explained by chronically ill smokers being less likely to participate. Fourth, in a recent MR study, smoking was associated with increased leukocytes, haematocrit, haemoglobin, and MCV, but not erythrocytes or platelets(Pedersen, *et al* 2019).

In GESUS and the meta-analysis, current smoking was associated with increased MCV (i.e. macrocytosis) and RDW, and likewise in the MR study, the smoking intensity was associated with increased MCV and RDW in current smokers. The findings for MCV can be explained by smoking and cotinine, a nicotine metabolite, also being associated with decreased levels of folate and cobalamin and increased levels of homocysteine(Haj Mouhamed, *et al* 2011). An increased RDW is associated with poor overall survival and poor cancer-specific survival in patients with cancer(Hu, *et al* 2017), particularly in patients with CML(Iriyama, *et al* 2015), but no studies have investigated RDW as a prognostic factor for MPN.

For CBC, the SMD allowed us to compare all cell types on the same scale, and the random effects statistic accounted for the high heterogeneity in study designs across studies. We investigated study characteristics such as smoking intensity (heavy vs. light), smoking status (current, ex, ever), sex, lab testing methodology, study size, year published, age distribution, or geographical region as culprits for heterogeneity. In the meta-analysis, we were not able to address pack-years, smoking duration, years since smoking cessation, or differences between filter vs. non-filter cigarettes, as reporting in studies were too diverse or lacking on these factors. A recent study showed that within ex-smokers, years since smoking cessation was inversely associated with white blood cells and the differentials, but values were still elevated compared to never smokers(Pedersen, *et al* 2019). Lastly, many unreported pre-analytical factors in the studies in the meta-analyses may have affected CBC such as seasonal and diurnal variations, fasting vs. non-fasting sampling, time since last cigarette smoked, position at blood draw (supine or sitting), altitude, physical activity, possible dormant infections, allergies, undiagnosed diseases, nutritional status, menstrual cycle and menopausal status.

For MPN, there was not enough information to address sex differences. Also, the description of MPN was not completely identical in the studies. Leal used international classification of diseases 9 (ICD-9)(Leal, *et al* 2014) and Pedersen ICD-8 and ICD-10(Pedersen, *et al* 2018). Sørensen recruited MPN patients from a single-institution based on WHO-5 MPN criteria(Lindholm Sorensen and Hasselbalch 2016). Kroll used ICD-10 and other classification systems and included myelodysplastic syndrome, CML, and other myeloproliferative disorders(Kroll, *et al* 2012), but it was not possible to filter out these diagnoses. Pasqualetti did not specify the MPN diagnosis or coding any further (Pasqualetti, *et al* 1997). None of the studies explicitly mentioned bone marrow pathology as part of the MPN diagnosis. Incidence rates (IR per 100,000 person-years) of MPN varied across the follow-up studies with an IR of 98 in Leal (Leal, *et al* 2014) with a median of 11 years of follow-up (~ 8.9 annually per 100,000), an IR of 13 in Pedersen with a mean of 6.8 years of follow-up (~ 1.9 annually per 100,000) (Pedersen, *et al* 2018), and an IR of 16 in Kroll with a mean of 10 years of follow-up (~1.6 annually per 100,000)(Kroll, *et al* 2012). In comparison, a recent meta-analysis reported an incidence rate of 2.6 annually per 100,000 for MPN(Titmarsh, *et al* 2014).

In conclusion, the association of smoking with hyperproliferation of blood cells and MPN were supported by the strength of the associations, the consistency across the large number of observational studies, the MR analysis, and the dose-response relationship. Furthermore, previous complementary studies support the mechanistic specificity of smoking with genotoxicity forgoing and being the driver of MPN. Future studies should elucidate when increased cell counts are merely secondary to smoking or reflect an underlying cancer. Our findings of an association between smoking and MPNs should alert clinicians that elevated cell counts in smokers may also reflect an underlying myeloproliferative cancer.

Acknowledgements

Search terms and creation of Endnote library with 18,516 references for the meta-analyses: Paul Bain, Countway Library of Medicine, 10 Shattuck street, Boston (paul_bain@hms.harvard.edu).

UK-biobank: extraction of data: Mathias N. Kjaergaard, Technology and Innovation, Terma A/S, Lystrup, Denmark (mnk@mnk.dk).

Authorship contributions

Conception of study design: NAJ, MB, MKL, BGN, SEB, YC, VHS, LK, HCH, CE. Study PIs for GESUS (SJ-113, SJ114, SJ452): CE, HCH, VHS, LK. Collection of participants for GESUS: CE, MKL. Meta-analysis: identification of articles, extraction of data, statistical coding, analyses, figures, tables: NAJ, CE. UK-biobank: statistical coding, analyses, figures, tables: ADK. Study PIs providing data for MPN meta-analyses from previously published studies: ALS, KMP, JST, HCH. Manuscript draft: NAJ and CE. Critical revision of manuscript and final approval: All authors.

Conflicts of interest disclosures

The authors do not have any conflicts of interest.

Funding

The Danish General Suburban Population Study was funded by the Region Zealand Foundation, Naestved Hospital Foundation, Johan and Lise Boserup Foundation, TrygFonden, Johannes Fog's Foundation, Region Zealand, Naestved Hospital, The National Board of Health, and the Local Government Denmark Foundation. Kasper M. Pedersen, Christina Ellervik, and Børge G. Nordestgaard was funded by Karen Elise Jensen's foundation.

Table 1. Characteristics of study designs on chronic cigarette smoking, complete blood count, and MPN

	Complete blood counts			MPN	All
	General population (A)	Meta-analysis** (B)	Mendelian Randomization (C)	Meta-analysis (D)	(B+C+D)
Study populations	GESUS	92	UK Biobank#	6	
Countries	Denmark	33	UK	5	
Publication years (yyyy-yyyy)	-	1974-2018	2018	1997-2018	
Age range, years	20-99	15-99	40-69	16-99	
Individuals, N*	11083	531741	350475	1425529	**2307745**
Men, N	5155	454003	162399	31377	**647779**
Women, N	5928	77252	188076	1391966	**1657294**
Current smokers, N	2217	214703	25348	267942	**507993**
Current heavy smokers, N	690	11881	-	5729	**17610**
Current light smokers, N	1371	22527	-	5088	**27615**
Ex-smokers, N	3933	90174	84456	380672	**555302**
Ever-smokers, N	6150	243969	109804	648614	**1002387**
Non-smokers, N	-	68840	-	671714	**740554**
Never-smokers, N	4933	153760	171500	-	**325260**

*Total individuals included were the sum of current smokers, ex-smokers, non-smokers, and never-smokers. If articles did not provide N for exposed/unexposed, total individuals were the number listed as total in the article.

**The meta-analysis of blood counts included GESUS. Never-smokers were life-long never-smokers. Non-smokers were non-current smokers. Men+women does not add up to total individuals, as some studies did not provide information on sex.

#UK biobank genotypes from Haplotype Reference Consortium (HRC) plus UK10K & 1000 Genomes reference panels as released by UK Biobank in March 2018. See text for details. Never smokers is approximately 49% of total, but the exact number for those with genotype+hematology was not publically available.

##Heavy smokers: >=20 cigarettes per day. Light-smokers <20 cigarettes per day.

GESUS: The Danish General Suburban Population Study. MPN: myeloproliferative neoplasms. UK: United Kingdom.

Figure Legend

Figure 1. Standardized adjusted mean differences in current smokers vs. never smokers – The Danish General Suburban Population Study

Figure 2. PRISMA flow chart. EMBASE: Excerpta Medica database. GESUS: The Danish General Suburban Population Study. WHO-GHL: World Health Organization Global Health Library.

Figure 3. Meta-analyses of complete blood cell counts - Current smokers vs non-smokers

Figure 4. Mendelian Randomization using *CHRNA3* genotype in UKBiobank – SD increase in complete blood cell counts per genetically one SD increase in cigarettes per day in current smokers. CHRNA: Cholinergic Receptor Nicotinic Alpha 3 Subunit (rs1051730 [A]). SMD: standardized mean difference (adjusted for sex, age, principal components).

Figure 5. Meta-analysis of Philadelphia-negative myeloproliferative neoplasms – Current smokers vs. non-smokers.

Figure 6. Meta-analyses of Philadelphia-negative chronic myeloproliferative neoplasms – By smoking status. P-Egger: publication bias p-value according to Eggers test. p-het: heterogeneity p-value. SMD: standardized mean difference.

References

Asgary, S., Naderi, G. & Ghannady, A. (2005) Effects of cigarette smoke, nicotine and cotinine on red blood cell hemolysis and their -SH capacity. *Exp Clin Cardiol,* **10,** 116-119.

Barbui, T., Thiele, J., Gisslinger, H., Kvasnicka, H.M., Vannucchi, A.M., Guglielmelli, P., Orazi, A. & Tefferi, A. (2018) The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion. *Blood Cancer J,* **8,** 15.

Baxter, E.J., Scott, L.M., Campbell, P.J., East, C., Fourouclas, N., Swanton, S., Vassiliou, G.S., Bench, A.J., Boyd, E.M., Curtin, N., Scott, M.A., Erber, W.N., Green, A.R. & Cancer Genome, P. (2005) Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet,* **365,** 1054-1061.

Bergholdt, H.K., Bathum, L., Kvetny, J., Rasmussen, D.B., Moldow, B., Hoeg, T., Jemec, G.B., Berner-Nielsen, H., Nordestgaard, B.G. & Ellervik, C. (2013) Study design, participation and characteristics of the Danish General Suburban Population Study. *Dan Med J,* **60,** A4693.

Billimoria, J.D., Pozner, H., Metselaar, B., Best, F.W. & James, D.C. (1975) Effect of cigarette smoking on lipids, lipoproteins, blood coagulation, fibrinolysis and cellular components of human blood. *Atherosclerosis,* **21,** 61-76.

Bjorn, M.E. & Hasselbalch, H.C. (2015) The Role of Reactive Oxygen Species in Myelofibrosis and Related Neoplasms. *Mediators Inflamm,* **2015,** 648090.

Brereton, M., McCafferty, R., Marsden, K., Kawai, Y., Etzell, J., Ermens, A. & international council for standardization in, h. (2016) Recommendation for standardization of haematology reporting units used in the extended blood count. *Int J Lab Hematol,* **38,** 472-482.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P. & Marchini, J. (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature,* **562,** 203-209.

Campbell, P.J., Scott, L.M., Buck, G., Wheatley, K., East, C.L., Marsden, J.T., Duffy, A., Boyd, E.M., Bench, A.J., Scott, M.A., Vassiliou, G.S., Milligan, D.W., Smith, S.R., Erber, W.N., Bareford, D., Wilkins, B.S., Reilly, J.T., Harrison, C.N., Green, A.R., United Kingdom Myeloproliferative Disorders Study, G., Medical Research Council Adult Leukaemia Working, P., Australasian, L. & Lymphoma, G. (2005) Definition of subtypes of essential thrombocythaemia and relation to polycythaemia vera based on JAK2 V617F mutation status: a prospective study. *Lancet,* **366,** 1945-1953.

Cordua, S., Kjaer, L., Skov, V., Pallisgaard, N., Hasselbalch, H.C. & Ellervik, C. (2019) Prevalence and phenotypes of JAK2 V617F and calreticulin mutations in a Danish general population. *Blood,* **134,** 469-479.

Dotevall, A., Kutti, J., Teger-Nilsson, A.C., Wadenvik, H. & Wilhelmsen, L. (1987) Platelet reactivity, fibrinogen and smoking. *Eur J Haematol,* **38,** 55-59.

Eisenga, M.F., Kieneker, L.M., Touw, D.J., Nolte, I.M., van der Meer, P., Huls, G., Gaillard, C. & Bakker, S.J.L. (2018) Active Smoking and Hematocrit and Fasting Circulating Erythropoietin Concentrations in the General Population. *Mayo Clin Proc,* **93,** 337-343.

Frederiksen, H., Farkas, D.K., Christiansen, C.F., Hasselbalch, H.C. & Sorensen, H.T. (2011) Chronic myeloproliferative neoplasms and subsequent cancer risk: a Danish population-based cohort study. *Blood,* **118,** 6515-6520.

Goldstein, M. (2008) Carbon monoxide poisoning. *J Emerg Nurs,* **34,** 538-542.

Haj Mouhamed, D., Ezzaher, A., Neffati, F., Douki, W. & Najjar, M.F. (2011) Effect of cigarette smoking on plasma homocysteine concentrations. *Clin Chem Lab Med,* **49,** 479-483.

Hasselbalch, H.C. (2012) Perspectives on chronic inflammation in essential thrombocythemia, polycythemia vera, and myelofibrosis: is chronic inflammation a trigger and driver of clonal evolution and development of accelerated atherosclerosis and second cancer? *Blood,* **119,** 3219-3225.

Hasselbalch, H.C. (2015) Smoking as a contributing factor for development of polycythemia vera and related neoplasms. *Leuk Res*.

Hasselbalch, H.C. & Bjorn, M.E. (2015) MPNs as Inflammatory Diseases: The Evidence, Consequences, and Perspectives. *Mediators Inflamm,* **2015,** 102476.

Hu, L., Li, M., Ding, Y., Pu, L., Liu, J., Xie, J., Cabanero, M., Li, J., Xiang, R. & Xiong, S. (2017) Prognostic value of RDW in cancers: a systematic review and meta-analysis. *Oncotarget,* **8,** 16027-16035.

Hultcrantz, M., Wilkes, S.R., Kristinsson, S.Y., Andersson, T.M., Derolf, A.R., Eloranta, S., Samuelsson, J., Landgren, O., Dickman, P.W., Lambert, P.C. & Bjorkholm, M. (2015) Risk and Cause of Death in Patients Diagnosed With Myeloproliferative Neoplasms in Sweden Between 1973 and 2005: A Population-Based Study. *J Clin Oncol,* **33,** 2288-2295.

Iriyama, N., Hatta, Y., Kobayashi, S., Uchino, Y., Miura, K., Kurita, D., Kodaira, H., Takahashi, H., Iizuka, Y., Inoue, M. & Takei, M. (2015) Higher Red Blood Cell Distribution Width Is an Adverse Prognostic Factor in Chronic-phase Chronic Myeloid Leukemia Patients Treated with Tyrosine Kinase Inhibitors. *Anticancer Res,* **35,** 5473-5478.

Kroll, M.E., Murphy, F., Pirie, K., Reeves, G.K., Green, J., Beral, V. & Million Women Study, C. (2012) Alcohol drinking, tobacco smoking and subtypes of haematological malignancy in the UK Million Women Study. *Br J Cancer,* **107,** 879-887.

Landtblom, A.R., Bower, H., Andersson, T.M., Dickman, P.W., Samuelsson, J., Bjorkholm, M., Kristinsson, S.Y. & Hultcrantz, M. (2018) Second malignancies in patients with myeloproliferative neoplasms: a population-based cohort study of 9379 patients. *Leukemia,* **32,** 2203-2210.

Lao, X.Q., Jiang, C.Q., Zhang, W.S., Adab, P., Lam, T.H., Cheng, K.K. & Thomas, G.N. (2009) Smoking, smoking cessation and inflammatory markers in older Chinese men: The Guangzhou Biobank Cohort Study. *Atherosclerosis,* **203,** 304-310.

Larsen, T.S., Pallisgaard, N., Moller, M.B. & Hasselbalch, H.C. (2007) The JAK2 V617F allele burden in essential thrombocythemia, polycythemia vera and primary myelofibrosis--impact on disease phenotype. *Eur J Haematol,* **79,** 508-515.

Leal, A.D., Thompson, C.A., Wang, A.H., Vierkant, R.A., Habermann, T.M., Ross, J.A., Mesa, R.A., Virnig, B.A. & Cerhan, J.R. (2014) Anthropometric, medical history and lifestyle risk factors for myeloproliferative neoplasms in the Iowa Women's Health Study cohort. *Int J Cancer,* **134,** 1741-1750.

Li, N., Zhou, H. & Tang, Q. (2017) Red Blood Cell Distribution Width: A Novel Predictive Indicator for Cardiovascular and Cerebrovascular Diseases. *Dis Markers,* **2017,** 7089493.

Lindholm Sorensen, A. & Hasselbalch, H.C. (2016) Smoking and philadelphia-negative chronic myeloproliferative neoplasms. *Eur J Haematol,* **97,** 63-69.

Liu, J.Z. & Tozzi, F. & Waterworth, D.M. & Pillai, S.G. & Muglia, P. & Middleton, L. & Berrettini, W. & Knouff, C.W. & Yuan, X. & Waeber, G. & Vollenweider, P. & Preisig, M. & Wareham, N.J. & Zhao, J.H. & Loos, R.J. & Barroso, I. & Khaw, K.T. & Grundy, S. & Barter, P. & Mahley, R. & Kesaniemi, A. & McPherson, R. & Vincent, J.B. & Strauss, J. & Kennedy, J.L. & Farmer, A. & McGuffin, P. & Day, R. & Matthews, K. & Bakke, P. & Gulsvik, A. & Lucae, S. & Ising, M. & Brueckl, T. & Horstmann, S. & Wichmann, H.E. & Rawal, R. & Dahmen, N. & Lamina, C. & Polasek, O. & Zgaga, L. & Huffman, J. & Campbell, S. & Kooner, J. & Chambers, J.C. & Burnett, M.S. & Devaney, J.M. & Pichard, A.D. & Kent, K.M. & Satler, L. & Lindsay, J.M. & Waksman, R. & Epstein, S. & Wilson, J.F. & Wild, S.H. & Campbell, H. & Vitart, V. & Reilly, M.P. & Li, M. & Qu, L. & Wilensky, R. & Matthai, W. & Hakonarson, H.H. & Rader, D.J. & Franke, A. & Wittig, M. & Schafer, A. & Uda, M. & Terracciano,

A. & Xiao, X. & Busonero, F. & Scheet, P. & Schlessinger, D. & St Clair, D. & Rujescu, D. & Abecasis, G.R. & Grabe, H.J. & Teumer, A. & Volzke, H. & Petersmann, A. & John, U. & Rudan, I. & Hayward, C. & Wright, A.F. & Kolcic, I. & Wright, B.J. & Thompson, J.R. & Balmforth, A.J. & Hall, A.S. & Samani, N.J. & Anderson, C.A. & Ahmad, T. & Mathew, C.G. & Parkes, M. & Satsangi, J. & Caulfield, M. & Munroe, P.B. & Farrall, M. & Dominiczak, A. & Worthington, J. & Thomson, W. & Eyre, S. & Barton, A. & Wellcome Trust Case Control, C. & Mooser, V. & Francks, C. & Marchini, J. (2010) Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet,* **42,** 436-440.

Lohler, J. & Wollenberg, B. (2019) Are electronic cigarettes a healthier alternative to conventional tobacco smoking? *Eur Arch Otorhinolaryngol,* **276,** 17-25.

Lowe, G.D.O., Lee, A.J., Rumley, A., Smith, W.C.S. & Tunstall-Pedoe, H. (1992) Epidemiology of haematocrit, white cell count, red cell aggregation and fibrinogen: The Glasgow MONICA study. *Clinical Hemorheology,* **12,** 535-544.

Nielsen, C., Birgens, H.S., Nordestgaard, B.G. & Bojesen, S.E. (2013) Diagnostic value of JAK2 V617F somatic mutation for myeloproliferative cancer in 49 488 individuals from the general population. *Br J Haematol,* **160,** 70-79.

Page, M.J., Shamseer, L. & Tricco, A.C. (2018) Registration of systematic reviews in PROSPERO: 30,000 records and counting. *Syst Rev,* **7,** 32.

Parry, H., Cohen, S., Schlarb, J.E., Tyrrell, D.A., Fisher, A., Russell, M.A. & Jarvis, M.J. (1997) Smoking, alcohol consumption, and leukocyte counts. *Am J Clin Pathol,* **107,** 64-67.

Pasqualetti, P., Festuccia, V., Acitelli, P., Collacciani, A., Giusti, A. & Casale, R. (1997) Tobacco smoking and risk of haematological malignancies in adults: a case-control study. *Br J Haematol,* **97,** 659-662.

Pedersen, K.M., Bak, M., Sorensen, A.L., Zwisler, A.D., Ellervik, C., Larsen, M.K., Hasselbalch, H.C. & Tolstrup, J.S. (2018) Smoking is associated with increased risk of myeloproliferative neoplasms: A general population-based cohort study. *Cancer Med,* **7,** 5796-5802.

Pedersen, K.M., Colak, Y., Ellervik, C., Hasselbalch, H.C., Bojesen, S.E. & Nordestgaard, B.G. (2019) Smoking and Increased White and Red Blood Cells. *Arterioscler Thromb Vasc Biol***,** ATVBAHA118312338.

Prieme, H., Loft, S., Klarlund, M., Gronbaek, K., Tonnesen, P. & Poulsen, H.E. (1998) Effect of smoking cessation on oxidative DNA modification estimated by 8-oxo-7,8-dihydro-2'-deoxyguanosine excretion. *Carcinogenesis,* **19,** 347-351.

Smith, M.R., Kinmonth, A.L., Luben, R.N., Bingham, S., Day, N.E., Wareham, N.J., Welch, A. & Khaw, K.T. (2003) Smoking status and differential white cell count in men and women in the EPIC-Norfolk population. *Atherosclerosis,* **169,** 331-337.

Spivak, J.L. (2017) Myeloproliferative Neoplasms. *N Engl J Med,* **376,** 2168-2181.

Titmarsh, G.J., Duncombe, A.S., McMullin, M.F., O'Rorke, M., Mesa, R., De Vocht, F., Horan, S., Fritschi, L., Clarke, M. & Anderson, L.A. (2014) How common are myeloproliferative neoplasms? A systematic review and meta-analysis. *Am J Hematol,* **89,** 581-587.

Tobacco & Genetics, C. (2010) Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet,* **42,** 441-447.

Wannamethee, S.G., Lowe, G.D., Shaper, A.G., Rumley, A., Lennon, L. & Whincup, P.H. (2005) The metabolic syndrome and insulin resistance: relationship to haemostatic and inflammatory markers in older non-diabetic men. *Atherosclerosis,* **181,** 101-108.

Weinberg, I., Borohovitz, A., Krichevsky, S., Perlman, R., Ben-Yehuda, A. & Ben-Yehuda, D. (2012) Janus Kinase V617F mutation in cigarette smokers. *Am J Hematol,* **87,** 5-8.