Title

Use of gender distribution in routine surveillance data to detect potential transmission of gastrointestinal infections among men who have sex with men in England

Authors

P Mook\textsuperscript{1, 2}, D Gardiner\textsuperscript{1a}, S Kanagarajah\textsuperscript{1a}, M Kerac\textsuperscript{1, 3, 4}, G Hughes\textsuperscript{5}, N Field\textsuperscript{5, 6}, N McCarthy\textsuperscript{1, 2, 7}, C Rawlings\textsuperscript{1}, I Simms\textsuperscript{5}, C Lane\textsuperscript{8}, P. D. Crook\textsuperscript{1b}

(2) Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK
(3) Department of Population Health, London School of Hygiene and Tropical Medicine, London, UK
(4) Leonard Cheshire Disability & Inclusive Development Centre, Department of Epidemiology & Public Health, University College London, UK
(5) HIV and STI Department, National Infection Service, Public Health England, London, UK
(6) Centre for Molecular Epidemiology and Translational Research, Institute for Global Health, University College London, London, UK
(7) National Institute Health Research (NIHR) Health Protection Research Unit in Gastrointestinal Infections, United Kingdom
(8) Gastrointestinal, Emerging and Zoonotic Infections Department, Public Health England, London, UK

\textsuperscript{a} These authors contributed equally
Corresponding author

Paul Crook, Consultant Epidemiologist, Field Epidemiology Service, Skipton House, 80 London Road, London, SE1 6LH, UK. Paul.crook@phe.gov.uk

Running head

Gender distribution of GI pathogens, England
Summary

Detecting gastrointestinal infection transmission among men who have sex with men (MSM) in England is complicated by a lack of routine sexual behavioural data. We investigated whether gender distributions might generate signals for increased transmission of gastrointestinal pathogens among MSM. We examined the percentage male of laboratory confirmed patient-episodes for patients with no known travel history for 10 gastrointestinal infections of public health interest in England between 2003 and 2013, stratified by age and region. An adult male excess was observed for *Shigella* spp. (annual maximum 71% male); most pronounced for those aged 25-49 years and living in London, Brighton and Manchester. An adult male excess was observed every year for *Entamoeba histolytica* (range 59.8-76.1% male), giardia (53.1-57.6%) and campylobacter (52.1-53.5%) and for a minority of years for hepatitis A (max. 69.8%) and typhoidal salmonella (max. 65.7%). This approach generated a signal for excess male episodes for six gastrointestinal pathogens, including a characterised outbreak of shigella among MSM. Stratified analyses by geography and age group were consistent with MSM transmission for shigella. Optimisation and routine application of this technique by public health authorities elsewhere might help identify potential gastrointestinal infection outbreaks due to sexual transmission among MSM, for further investigation.
Introduction

Gastrointestinal (GI) pathogens including shigella [1, 2], campylobacter [3, 4], vero cytotoxin-producing *Escherichia coli* (VTEC) [5], giardia [6, 7], *Salmonella enterica* serotype Typhi [8], cryptosporidium [9, 10], hepatitis A [11-14] and enteric protozoans, including *Entamoeba histolytica* [15], can be spread through sexual contact, most commonly among men who have sex with men (MSM)[16]. Risk of infection is likely influenced by sexual practices, infectivity and HIV status [16, 17].

In England, sexual history is not routinely collected for most cases of laboratory confirmed gastrointestinal infections reported to Public Health England (PHE). In the absence of this information, there is the risk that increases or outbreaks among MSM may go undetected. However, increased transmission of GI infections among MSM might produce a detectable signal in the gender distribution among cases reported in routine surveillance data. Gender ratios have been applied previously to demonstrate that HIV infection was transmitted predominantly between heterosexuals in Africa in the 1980’s [18] and as a surrogate marker for MSM activity in Atlanta, USA [6].

A rise in *Shigella flexneri* 3a in England has been described among men without a known history of recent travel from 2009. Follow-up of a sub-set of cases suggested that faecal-oral transmission occurred during sexual contact between MSM, many of whom were HIV positive and reported high numbers of regular and casual partners, chemsex (engaging in sexual activities while under the influence of drugs) and meeting sex partners and locating sex parties through social and sexual networking [19-21]. The increase in shigellosis has coincided with increasing trends in other
sexually transmitted infections among MSM, including lymphogranuloma venereum, which have been associated with similar risk behaviours [22-24].

We investigated a range of GI pathogens over an 11 year period to explore whether the gender distribution would provide signals of potential sexual transmission, including known outbreaks among MSM.

**Methods**

**Study population**

The study population was residents of England aged 0-65 years, with a laboratory confirmed diagnosis of one of 10 GI pathogens of public health significance reported to PHE with a specimen date between 1 January 2003 and 31 December 2013, known gender and no known history of recent travel. The GI pathogens included were: campylobacter, cryptosporidium, *Entamoeba histolytica*, *Giardia* spp., hepatitis A, norovirus, typhoidal and non-typhoidal salmonellas, shigella, and vero cytotoxin-producing *Escherichia coli* (VTEC).

**Data source**

Data were extracted from Labbase, the national laboratory reporting system for England until March 2015, which stored data submitted from laboratories throughout England, Wales, Northern Ireland and the Channel Islands. For some pathogens - including shigella and salmonella - further typing is performed on a subset of samples submitted to national reference laboratories and these results supplement the Labbase data.
Data analysis

Confirmed laboratory diagnoses for the same pathogen in the same person within a 2 week period (26 weeks for salmonella) of the earliest specimen were de-duplicated and considered as one case-episode, as per established standards used to de-duplicate Labbase data. This was performed to reduce double counting individuals with persisting gastrointestinal infection who have multiple samples taken (e.g. for establishing clearance).

Analyses were restricted to cases resident in England, based on either the postcode of the case, general practice or reporting laboratory, in priority order. Cases were included if the date of their earliest specimen for a given episode was received by the reporting laboratory between 01 January 2003 and 31 December 2013.

Cases with any known recent foreign travel were excluded from the study. Cases were excluded if the laboratory report form noted the case having had any travel to a foreign country or listed a travel destination which was outside of the UK prior to symptom onset.

In the primary analysis for each pathogen we examined total and annual percentage male and male-to-female ratios for those aged 16-65 years over the study period. We made an assumption that in the absence of transmission among MSM by sexual contact we would expect 50% of cases to be male, with a 1:1 male-to-female ratio, for each pathogen. Binomial exact confidence intervals were calculated for the percentage male and a positive signal generated if the lower confidence interval was above 50%. We also reviewed data to note where male-to-female ratios rose above
two, as an arbitrary cut-off as being suggestive of MSM transmission. Chi-squared
tests for linear trend were applied to assess change in the gender distribution over
the study period at the 5% level for each pathogen.

For pathogens with a signal, secondary analyses were undertaken. Comparative
annual analyses were also conducted by age group (less than 16 years, 16-24
years, 25-49 years, 50-65 years) and areas with relatively high MSM populations in
England (London, Brighton and Manchester [25]); termed high risk areas) versus
elsewhere in England (termed low risk areas) to explore whether the percentage
male and male-to-female ratio varied with expected differences in the MSM
population distribution by age group and region. Further investigation of percentage
male and male-to-female ratios were conducted for shigella species and phage-
types.

Results

Over the study period, 529315 GI infection cases were reported in those aged 16-65
years (Table 1). Of these, 25192 cases (4.8%) were excluded as they reported a
travel history, leaving 504123 cases aged 16-65 included in the study with no or
unknown travel history.

The percentage of cases aged 16-65 years excluded as travel related (4.8%) ranged
by pathogen from 0.1% for norovirus; (n=18) to 56.9% for typhoidal salmonella
(n=1959) (Table 1). The percentage of cases classified as having an unknown travel
history (81%, n=428669) ranged from 36.2% (1247) for typhoidal salmonella to
92.2% (n=3358) for hepatitis A. The number of cases included ranged from 1483 for typhoidal salmonella to 382641 for campylobacter.

Over the whole 11 year study period combined, a positive signal (a male percentage with a lower confidence interval higher than 50%) was observed in adults (16-65 years) for six out of the 10 pathogens studied (Entamoeba histolytica 68.3% male, hepatitis A 61.1%, typhoidal salmonella 55.4%, giardia 54.8%, campylobacter 52.8%, shigella spp. 51.3%) and in three out of four of the shigella species and phage-types studied (Table 2). The largest number of excess adult male cases was observed for campylobacter (21649) and the highest adult male percentage was observed for Shigella flexneri PT3a cases (89.3%). An excess of females was observed among adults for cryptosporidium, norovirus, non-typhoidal salmonella, VTEC and Shigella sonnei. For three of the six pathogens with a male excess in adults (Shigella spp., campylobacter and giardia), there was a corresponding male excess in children, which was not observed for Entamoeba histolytica, typhoidal salmonella, and hepatitis A. For Shigella spp. as a whole, a positive signal was seen in high risk areas (62.5% male) but not in low risk areas (44.8%) or in children (52.1%). For Shigella flexneri, there was a positive signal for Shigella flexneri in adults, in both high (73.7% male) and low risk areas (53.2%), but not in children. For Shigella sonnei there was a positive signal in high risk areas (60% male) and in children (52.7%) but not in low risk areas (41.9%).
When reviewing individual years, no positive signals were observed in any of the 11 study years for cryptosporidium, norovirus, non-typhoidal salmonella, or VTEC (Table 3, Figure 1).

A positive signal for *Shigella* spp. in adults first occurred in 2011 with a subsequent significant rise (max. 71%, max m:f ratio 2.5), *Shigella flexneri* from 2010 onwards (max. 84.1%, max m:f ratio 5.3) and *Shigella sonnei* from 2012 (max. 63.5%, max m:f ratio 1.7) (Figures 1 and 2, Tables 3 and 4). A stronger signal in adults was seen for *Shigella flexneri* PT2a (max 84.1%, max m:f ratio 5.3) and *Shigella flexneri* PT3a (max 100%, max m:f ratio ∞). Annual data by age group and location for *Shigella* spp. showed male exceedances being higher, earlier and more frequent in those aged 25-49 years than other age groups, and for cases in high risk areas compared to low risk areas (Supplementary material S1). No individual year male excess was observed for *Shigella* spp. in children.

For *Entamoeba histolytica* a positive signal in adults was observed in all of the 11 study years for all adults (max. 76.1%, max m:f ratio 3.2), in 10 study years for those aged 25-49 years, in three years for those aged 50-65 years, in no years for 16-24 year olds and once for children (Figure 1, Supplementary material S2). A positive signal was observed in nine study years for both high risk and low risk areas.

A positive signal in adults was observed for hepatitis A in six out of 11 study years (max. 69.8%, with a significant falling linear trend at the 5% level) with signals were
more frequently observed in cases aged 25-49 years (six years), than those aged 16-24 years (four) and children (one), and from low risk (seven) compared with high risk areas (four) (Figure 1, Table 3, Supplementary material S3).

For typhoidal salmonella a positive signal in adults was observed in three of the 11 study years (max. 65.7% with a significant falling trend) and signals were more frequently seen in cases aged 25-49 years (three years), than those aged 16-24 years (none), children (one), and in cases from high risk areas (three) than low risk areas (one) (Figure 1, Table 3, Supplementary material S4).

Positive signals were observed in adults for every year of the study period for campylobacter (maximum 53.5%, max m:f ratio 1.1) (Figure 1, Table 3). A signal was observed in nearly every year for each age group and every year for low risk areas and in nine study years in high risk areas. The percentage male was higher in children than in adults.

For giardia, a positive signal in adults was observed for each of the 11 study years (max 57.6%, max m:f ratio 1.4) (Figure 1, Table 3). A signal was more frequently observed for adults aged 25-49 years and children (all years), compared to 16-24 year olds (two years) and 50-65 year olds (one), and in low risk areas (all years) compared to high risk areas (ten).
For adults aged 15-65 years a male-to-female ratio of cases above an arbitrary cut-off of two was only observed for *Entamoeba histolytica*, hepatitis A and *Shigella* spp. for England as a whole (Table 3). For hepatitis A and *Entamoeba histolytica* there were no occasions where the male-to-female ratio was above two and where the adult male percentage did not also provide a signal. There was a mixed picture for shigella. For *Shigella flexneri* the male-to-female ratio rose above two in 2012, later than the first adult male percentage signal (2010) (Table 4). For *Shigella flexneri* PT2a and PT3a the male-to-female ratio rose above two prior to the adult male percentage signal (albeit with small numbers of cases). For *Shigella sonnei*, the male-to-female ratio did not rise above two, while the percentage adult male did signal from 2012.

**Discussion**

We have applied analysis of male percentage and male-to-female ratios to surveillance data to identify excess gastrointestinal infections among males. This approach generated positive signals for excess male episodes for a period with a well-characterised increase in shigella among MSM. Positive signals were also observed for campylobacter, *Entamoeba histolytica*, giardia, typhoidal salmonella and hepatitis A. No signals were detected for cryptosporidium, norovirus, VTEC or non-typhoidal salmonella spp. Our analysis suggests that routinely collected national surveillance data can be used to help assess the potential contribution of sexual exposure to the transmission of GI pathogens and to detect emerging outbreaks in MSM. Male excess analysis should be seen as hypothesis generating and any signal
detected needs further thorough case level investigation to confirm sexual transmission among MSM.

When using male excess signals among adults to highlight potential MSM transmission, other factors that may result in a male excess need to be considered, including an excess of males in the population, a reduction in female cases, changes in testing and reporting practice, gender specific health seeking behaviour, or random variation. Furthermore, other gender disparities in behaviour such as travel, childcare, injecting drug use and food consumption may influence exposure to GI pathogens such that comparison with a 1:1 male-to-female ratio may not be appropriate. However, adult females are more likely to present to primary care than adult males likely resulting in a testing bias that would tend to lower the male-to-female ratio [26]. To strengthen the hypothesis that a signal truly represents an increase in MSM sexual transmission, one would expect the signal to be greater in adult males aged 25-49 years compared to other age groups, especially children. A greater signal in high risk areas may also support the hypothesis for widespread MSM sexual transmission contributing to overall cases.

While we analysed a large national 11 year dataset there were limitations. Recent travel history was poorly reported. Inclusion of cases with undocumented travel history might lead to misclassification of travel associated infection as domestically acquired. The extent and direction of potential bias is difficult to determine and will vary by pathogen, as it is dependent on a number of factors, including the proportion of cases reporting travel abroad, the proportion of missing travel information, the likelihood of sexual transmission while abroad and gender differences in travel
abroad. International Passenger Survey (IPS) data for 2013 indicates that UK adult males are more likely to travel internationally than females [27] and therefore misclassification may lead to an increase in the observed male percentage. However, misclassification may result in dilution of effect if a high proportion of unknown travel history cases are both likely to be travel related and associated with food- or waterborne infection and such exposure is independent of gender. The numbers presented here likely underestimate the true counts as not all cases in the community present to health care or provide specimens (and the proportion who do differ by pathogen) [28].

Use of the male-to-female ratio as a marker for MSM activity has been applied previously and its usefulness described [6, 18]. Retrospective application of our method detected the known increase in *S. flexneri* PT 2a and 3a and *S. sonnei* among MSM [19, 29], and an outbreak of hepatitis A among MSM in 2004 [30]. However, other years in which there was a positive signal for hepatitis A coincided with documented outbreaks among people who inject drugs [31, 32], who are over-represented by young adults and men [33], and the Orthodox Jewish Community [34]. Overall, then, our refined approach using age and geographical strata has validated the use of male-to-female ratio for highlighting potential MSM transmission.

The age and geographical distributions of excess male signals for *Shigella* spp. were consistent with MSM-mediated transmission. Signals were more frequently observed in areas with large MSM populations and high rates of other sexually transmitted infections such as London, Brighton and Manchester, and no signal was seen among children. Application of the excess male percentage method to distinct
species and age groups appeared to provide more discriminatory power. A signal was detected one year earlier both for *Shigella* spp. in adults aged 25-49 years when compared to all adults, and for *Shigella flexneri* when compared to *Shigella* spp.. It is possible that more signals for other organisms would have been observed with further discriminatory typing data.

Among other pathogens with positive signals for male excess cases, *Entamoeba histolytica* and hepatitis A showed the strongest indication of likely MSM transmission, with more signals in adults than in children, and for *Entamoeba histolytica*, stronger signals in high risk compared to low risk areas. Transmission among MSM has been well described for both these organisms [11-13, 15]. The episodic positive signal for typhoidal salmonella may reflect the effect of bias arising from incomplete travel information as the majority of cases included in the study with no unknown travel were likely to have travelled (57% of typhoidal salmonella cases identified were excluded due to known travel and only 7% were known not to have travelled).

The consistent slight male excess of campylobacter and giardia are of interest. The burden of campylobacter in England is much greater than other GI infections and therefore we found a very high total excess number of adult males over the study period. Transmission among MSM have been described for both campylobacter [3, 4] and giardia [6, 7], however, the finding for both organisms that the excess is consistent in individual years, in both children and adults, and in low as well as high risk areas may mean that gender factors other than sexual transmission are important in explaining the male excess among adults in this study. Campylobacter
infections in England and Wales have previously been reported to be more common in men up until 15 years and thereafter more common in women [35], and in a study among English residents of Pakistani origin with campylobacter infection, a higher proportion were males [36].

We analysed both the percentage male (applying confidence intervals) and the male-to-female ratio. The sensitivity of using an arbitrary male-to-female ratio threshold e.g. two, to trigger action is greatly influenced by the number of cases due to other transmission routes for each pathogen. In our study the number of campylobacter cases was much higher than shigella cases and for a given number of cases due to sexual transmission among MSM, the male-to-female ratio would be higher for shigella than campylobacter. In addition, the baseline male-to-female ratio in the absence of sexual transmission appears to differ by pathogen e.g. the cryptosporidium male-to-female ratio consistently remained below one. Given these factors, pathogen specific threshold ratios which trigger action could be refined, modelled on historical data.

In our study we excluded all known travel related cases, prioritising identifying domestic sexual transmission. However, MSM may acquire infections due to sexual activity abroad and this method could be refined for each organism to include travel to destinations where it was considered that sexual transmission may be more likely than other transmission routes e.g. via food or water.

Monitoring gender differences can be seen as a supplementary approach to existing measures. Traditional outbreak detection methods e.g. weekly statistical exceedance
of total counts compared to expected values determined from recent years, may
detect outbreaks due to MSM transmission, when subsequent descriptive
epidemiology points to an excess of adult males. However, there are a number of
reasons why MSM outbreaks may be more likely to go undetected than food-borne
outbreaks. The person to person nature of MSM transmission (as opposed to point
source food-borne outbreaks) may result in a rising tide of cases over many months
and therefore may be less to likely trigger weekly exceedances. Furthermore, food-
borne outbreaks may be detected due to sick diners recognising a common event
and approaching authorities. In contrast, MSM may incorrectly ascribe their illness
due to food poisoning and therefore it is less likely that links to sexual transmission
events will be reported. A foodborne outbreak may also result in subsequent MSM
transmission and so periodic review of gender difference for known prolonged
outbreaks may help identify a change in transmission route.

There is growing evidence that transmission of GI pathogens among MSM is
becoming a public health concern globally, especially among HIV-positive MSM
reporting high risk sexual and drug use behaviours [16, 17, 37-39]. In England,
outbreaks of *Shigella flexneri*, have been associated with high rates of hospitalisation
and reports of bacteraemia [20, 21]. More recently, there has been an increase in the
male-to-female ratio of *Shigella sonnei* cases [29], characterised clusters of
extended spectrum beta-lactamase producing *Shigella sonnei* [40], vero cytotoxin-
producing *Escherichia coli* O117:H7 among (mostly HIV-positive) MSM [5] and an
international outbreak of hepatitis A [14]. Simple methods to improve detection of
outbreaks of GI pathogens that may lead to severe health outcomes are therefore
needed. In the absence of routinely collected information on sexual orientation,
routine application of this rapid method, refined to use pathogen specific male-to-female ratio or percentage male thresholds at the most granular typing discrimination available, might be a useful tool to alert public health authorities to potential GI infection outbreaks associated with sexual contact among MSM and afford earlier health promotion and interventions. We recommend that this approach be used by other countries to detect excess male cases and prompt further investigations to assess whether sexual transmission of GI pathogens among MSM warrants public health action. Surveillance could be further improved by introducing simple questions on recent sexual exposure in routine questionnaires for GI pathogens.

Acknowledgements

The research was in part funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at University of Liverpool in partnership with Public Health England (PHE), in collaboration with University of East Anglia, University of Oxford and the Institute of Food Research. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

Funding

This work was supported by the National Institute for Health Research (HPRU-2013-10038).

Conflicts of interest
References


