Electronic Journal of Medical Research
https://respublishers.com/journals/ejmr
ISSN: 3106-1362/3106-1370

Research article

The Limits of a Reporting Baseline: Sex Differences in Adverse Event Reporting for Tumour Necrosis Factor Blockers, GLP-1 Receptor Agonists and Proton Pump Inhibitors Across 1,753,642 Reports

Nabila Ghaffar [1,*] and Kinza Tahir [2]

[1] Women Medical Officer, IRHC Hafizwala, Tehsil Piplan, District Mianwali, Punjab, Pakistan
[2] Department of Medical Education, Allama Iqbal Medical College, Lahore, Punjab, Pakistan

*Correspondence: nabilaghafar03@gmail.com

Article information:
Received: 30-04-2026
Revised: 15-05-2026
Accepted: 15-05-2026
Published: 16-05-2026

Academic Editor:
Dr. Naveed Ahmed

Keywords:
Pharmacovigilance
Tumor Necrosis Factor Inhibitors
Agonists
Proton Pump Inhibitors
Alopecia

Abstract: We have previously shown, across seven drug classes, that male adverse event reports are systematically more serious and that the reporting odds ratio (ROR) for death clusters near 0.63 irrespective of pharmacology. Whether this baseline is universal, and what would break it, was unknown. Case/non-case disproportionality analysis of the FDA Adverse Event Reporting System via openFDA database. All reports listing a tumour necrosis factor (TNF) blocker, GLP-1 receptor agonist or proton pump inhibitor (PPI) with a recorded patient sex were included. RORs for women versus men were calculated within each class for 34 pre-specified Preferred Terms, with Benjamini-Hochberg control of the false discovery rate. 1,753,642 reports were analysed. Male reports were more often serious in all three classes, making ten of ten. The death ROR was 0.59 for TNF blockers and 0.63 for PPIs, within the established range, but 0.38 for GLP-1 agonists, 7.4 SD below the nine-class mean of 0.632. GLP-1 reports were also far less often serious than any other class (27.3% in women). Alopecia gave the largest effects in the programme and was near-identical in two unrelated classes: ROR 8.34 (95% CI 7.74-8.98) for TNF blockers and 8.15 (7.13-9.30) for GLP-1 agonists. PPI outcomes divided between gastrointestinal symptoms in women and renal injury in men. The severity baseline is robust but not universal. It fails where the reporting population differs markedly, which bounds its generality and clarifies its cause. The near-identical alopecia signal across unrelated classes suggests differential reporting of salient events.

Article citation: Nabila, G., & Tahir, K. The Limits of a Reporting Baseline: Sex Differences in Adverse Event Reporting for Tumour Necrosis Factor Blockers, GLP-1 Receptor Agonists and Proton Pump Inhibitors Across 1,753,642 Reports. Electron J Med Res. 2026. 2(2): 54-60. https://doi.org/10.64813/ejmr.2026.089.

Introduction

Spontaneous reporting systems record what is reported rather than what occurs. In two preceding analyses using a common within-class design, we found that this distinction has a measurable and remarkably stable structure [1]. Across seven pharmacologically unrelated drug classes, male reports were more often flagged as serious than female reports, and the reporting odds ratio for death clustered between 0.60 and 0.70 with a standard deviation of 0.035 [2, 3].

We proposed that this constitutes a baseline against which sex-stratified disproportionality estimates should be read, and that an ROR near 0.63 for a severe outcome is the expected value rather than a signal. A claim of that kind carries an obligation [4]. If the baseline is a general property of reporting, it should appear in classes chosen for their difference from those already studied; and if it is not general, the circumstances under which it fails should be identifiable and should illuminate its cause [5, 6]. We therefore selected three classes that differ from the preceding seven and from one another in indication, route, patient population and reporting context: TNF

Copyright © 2026 Nabila, G., & Tahir, K. This work is published and Licensed by RES Publishers.
The full terms of this license are available at https://www.respublishers.com/terms_conditions.php.

blockers, given by injection for chronic inflammatory disease; GLP-1 receptor agonists, increasingly used for weight reduction in people who are not otherwise unwell; and proton pump inhibitors, largely oral, often over-the-counter, and used across the whole adult age range. The GLP-1 class was included specifically because its reporting population seemed least like the others.

Material and Methods

Data source and cohort

The FDA Adverse Event Reporting System was accessed through the public openFDA application programming interface, using the (20,692,690 reports). Reports were identified by the Established Pharmacologic Class terms 'Tumor Necrosis Factor Blocker', 'GLP-1 Receptor Agonist' and 'Proton Pump Inhibitor' applied to any listed drug irrespective of suspect status, and were included where patient sex was recorded as male or female. No ethical approval was required; the database is public and de-identified.

Outcome selection

For each class we pre-specified Preferred Terms spanning its recognised adverse effect profile, including both symptomatic and organ-level outcomes, with death as a common comparator. Terms whose count fell below the retrievable frequency range in either sex stratum were excluded rather than estimated, because the public interface returns only a bound in that case. Three TNF blocker terms of interest were lost this way, including tuberculosis and lymphoma, where the male count was retrievable but the female count was not; the direction of those associations is therefore known but their magnitude is not, and they are omitted from all analyses.

Statistical analysis

Analyses were conducted within each class so that drug exposure was held constant and sex was the exposure of interest. For each Preferred Term a 2 x 2 table compared reports containing that term against all other reports of the same class, separately by sex. RORs were calculated with 95% confidence intervals from the standard error of the log odds ratio [3, 7]. Thirty-four comparisons were made and the false discovery rate was controlled by the Benjamini-Hochberg procedure across all of them [8].

The death ROR for each class was compared against the distribution established across the seven classes analysed previously, expressed as a standardised deviation from that mean. Proportions were compared by chi-squared test with continuity correction. Analyses used Python 3 with SciPy. The study is reported in accordance with the STROBE statement [9].

Results

Cohort

1,753,642 reports with recorded sex were analysed: 834,940 TNF blocker reports (596,880 women, 238,060 men), 392,998 GLP-1 receptor agonist reports (263,989 women, 129,009 men) and 525,704 PPI reports (311,998 women, 213,706 men).

Table 1. Severity baseline across ten drug classes analysed under one design.

Drug class	Reports, women	Reports, men	Serious, women	Serious, men	ROR, death
TNF blocker	596,880	238,060	48.4%	56.2%	0.59
GLP-1 receptor agonist	263,989	129,009	27.3%	34.9%	0.38
Proton pump inhibitor	311,998	213,706	74.9%	81.5%	0.63
Opioid agonist*	168,930	139,460	70.8%	77.9%	0.70
Atypical antipsychotic*	200,589	201,115	78.5%	80.8%	0.66
Antiepileptic agent*	118,132	73,783	73.4%	80.2%	0.65
NSAID*	698,719	484,808	68.9%	75.1%	0.63
Corticosteroid*	651,361	491,276	75.5%	82.0%	0.61
HMG-CoA reductase inhibitor*	180,392	168,464	68.7%	75.6%	0.61
Benzodiazepine*	407,673	230,043	70.9%	79.8%	0.60

All sex differences in the proportion of reports flagged serious are significant at $p < 0.001$. ROR, reporting odds ratio for death, women versus men, within class. *Analysed previously under the same design. Excluding GLP-1 receptor agonists, the death ROR across nine classes ranges 0.59 to 0.70 (mean 0.632, SD 0.034); the GLP-1 value of 0.38 lies 7.4 standard deviations below that mean.

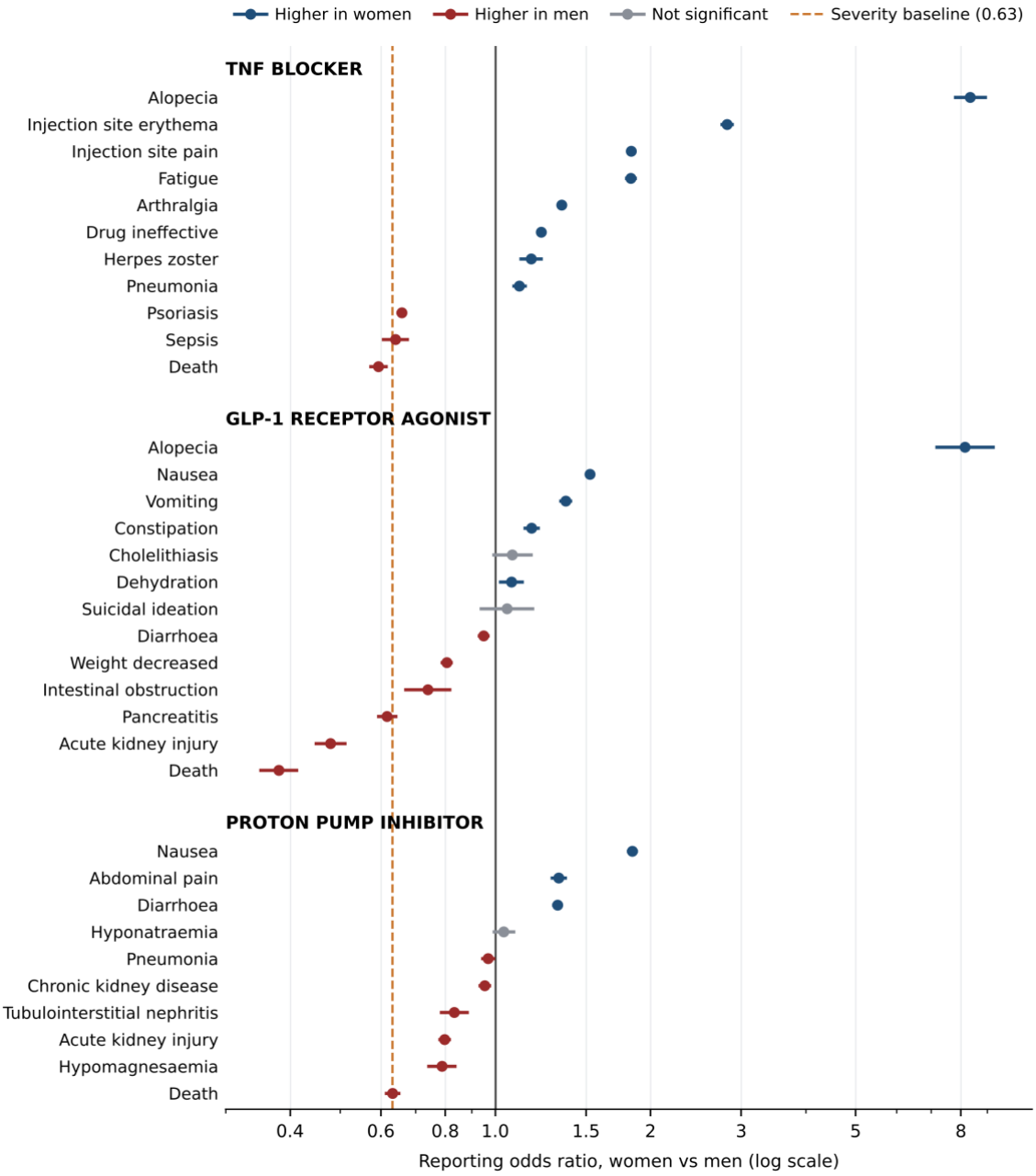


Figure 1. Reporting odds ratios for women versus men across three drug classes. Points are reporting odds ratios with 95% confidence intervals on a logarithmic scale. Values right of the solid line are reported disproportionately in women; left, in men. The dashed line marks the severity baseline of 0.63 derived from nine drug classes. Note the near-identical alopecia estimates in two pharmacologically unrelated classes.

The baseline holds, then fails

Male reports were more often flagged as serious in all three classes: 56.2% versus 48.4% for TNF blockers, 34.9% versus 27.3% for GLP-1 agonists and 81.5% versus 74.9% for PPIs (all $p < 0.001$). With the seven classes analysed previously, this direction has now been observed in ten of ten **(Figure 2B)**.

Table 2. Reporting odds ratios for women versus men, by drug class.

MedDRA Preferred Term	Women, n	Men, n	ROR (95% CI)	q
TNF BLOCKER				
Alopecia	15,124	740	8.34 (7.74–8.98)	$< 10^{-300}$
Injection site erythema	33,732	4,960	2.82 (2.73–2.90)	$< 10^{-300}$
Injection site pain	45,111	10,160	1.83 (1.79–1.87)	$< 10^{-300}$
Fatigue	32,163	7,180	1.83 (1.78–1.88)	$< 10^{-300}$
Arthralgia	42,990	12,985	1.35 (1.32–1.37)	3.6×10^{-182}
Drug ineffective	73,944	24,598	1.23 (1.21–1.25)	1.1×10^{-151}
Herpes zoster	5,517	1,878	1.17 (1.11–1.24)	3.6×10^{-9}
Pneumonia	14,233	5,112	1.11 (1.08–1.15)	1.2×10^{-10}
Psoriasis	18,431	10,988	0.66 (0.64–0.67)	2.3×10^{-255}
Sepsis	2,733	1,699	0.64 (0.60–0.68)	1.8×10^{-47}
Death	5,716	3,819	0.59 (0.57–0.62)	1.7×10^{-138}
GLP-1 RECEPTOR AGONIST				
Alopecia	3,818	232	8.15 (7.13–9.30)	2.4×10^{-297}
Nausea	39,933	13,502	1.52 (1.49–1.56)	$< 10^{-300}$
Vomiting	18,922	6,890	1.37 (1.33–1.41)	4.2×10^{-104}
Constipation	10,328	4,319	1.18 (1.13–1.22)	2.9×10^{-18}
Cholelithiasis	1,490	675	1.08 (0.99–1.18)	0.11
Dehydration	3,978	1,812	1.07 (1.02–1.14)	0.015
Suicidal ideation	823	382	1.05 (0.93–1.19)	0.42
Diarrhoea	17,185	8,813	0.95 (0.92–0.98)	1.7×10^{-4}
Weight decreased	14,567	8,729	0.80 (0.78–0.83)	2.7×10^{-54}
Intestinal obstruction	877	579	0.74 (0.67–0.82)	2.5×10^{-8}
Pancreatitis	4,170	3,272	0.62 (0.59–0.65)	2.1×10^{-94}
Acute kidney injury	1,488	1,509	0.48 (0.45–0.51)	5.5×10^{-93}
Death	901	1,152	0.38 (0.35–0.41)	9.2×10^{-112}
PROTON PUMP INHIBITOR				
Nausea	23,257	8,942	1.84 (1.80–1.89)	$< 10^{-300}$
Abdominal pain	8,751	4,548	1.33 (1.28–1.38)	7.3×10^{-53}
Diarrhoea	20,683	10,910	1.32 (1.29–1.35)	5.8×10^{-115}
Hyponatraemia	3,877	2,558	1.04 (0.99–1.09)	0.15
Pneumonia	9,518	6,730	0.97 (0.94–1.00)	0.048
Chronic kidney disease	12,615	9,041	0.95 (0.93–0.98)	9.5×10^{-4}
Tubulointerstitial nephritis	2,068	1,699	0.83 (0.78–0.89)	3.3×10^{-8}
Acute kidney injury	10,958	9,338	0.80 (0.77–0.82)	3.1×10^{-56}
Hypomagnesaemia	1,913	1,661	0.79 (0.74–0.84)	1.9×10^{-12}
Death	6,371	6,827	0.63 (0.61–0.65)	3.9×10^{-151}

ROR, reporting odds ratio; values above 1 indicate disproportionate reporting in women. Denominators are the class totals in Table 1. q values are Benjamini-Hochberg adjusted across all 34 comparisons. Terms are ordered by ROR within each class.

The death ROR was 0.59 for TNF blockers and 0.63 for PPIs, both squarely within the established range. For GLP-1 receptor agonists, it was 0.38 (95% CI 0.35-0.41), which is 7.4 standard deviations below the mean of 0.632 derived from the other nine classes **(Figure 2A)**. This is not a marginal departure and cannot be attributed to sampling variation.

The same class is also an outlier in absolute seriousness. Only 27.3% of GLP-1 reports concerning women were flagged serious, against 48.4% to 78.5% in every other class examined. The two observations are consistent with one another: a reporting population dominated by non-serious events will generate a different ratio of fatal to non-fatal reports, and the baseline is therefore a property of a reporting population rather than a constant of the database.

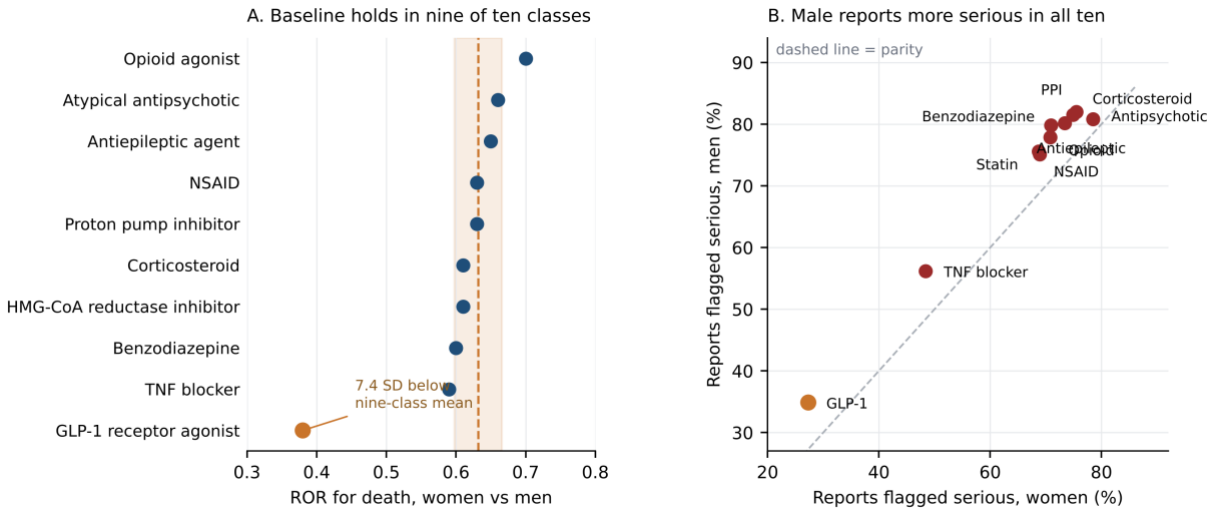


Figure 2. The severity baseline across ten drug classes. **(A)** Reporting odds ratio for death, women versus men, in each of the ten drug classes analysed under this design. The shaded band spans one standard deviation about the nine-class mean of 0.632, excluding GLP-1 receptor agonists. **(B)** Proportion of reports flagged serious, men against women; every class lies above parity, but GLP-1 receptor agonists and TNF blockers occupy a markedly lower absolute range than the rest.

Tumour necrosis factor blockers

Injection-related and cutaneous outcomes were strongly female-predominant: alopecia (ROR 8.34, 95% CI 7.74-8.98), injection site erythema (2.82, 2.73-2.90), injection site pain (1.83, 1.79-1.87) and fatigue (1.83, 1.78-1.88). Infection outcomes diverged by severity: herpes zoster (1.17, 1.11-1.24) and pneumonia (1.11, 1.08-1.15) were modestly female-predominant, whereas sepsis was male-predominant (0.64, 0.60-0.68), as was death (0.59, 0.57-0.62). Paradoxical psoriasis was male-predominant (0.66, 0.64-0.67).

Sepsis at 0.64 and death at 0.59 are indistinguishable from the severity baseline and should not be read as sex-specific signals. The infection gradient within this class nonetheless mirrors the symptom-versus-lesion structure reported previously for statins and NSAIDs.

GLP-1 receptor agonists

Gastrointestinal symptoms were female-predominant: nausea (ROR 1.52, 95% CI 1.49-1.56), vomiting (1.37, 1.33-1.41) and constipation (1.18, 1.13-1.22). Alopecia was again extreme at 8.15 (7.13-9.30). Structural and metabolic complications ran the other way: pancreatitis (0.62, 0.59-0.65), acute kidney injury (0.48, 0.45-0.51) and intestinal obstruction (0.74, 0.67-0.82). Weight decrease was male-predominant (0.80, 0.78-0.83). Cholelithiasis (1.08, 0.99-1.18) and suicidal ideation (1.05, 0.93-1.19) showed no sex difference.

Acute kidney injury at 0.48 and death at 0.38 both lie below even the lowered baseline for this class and are the estimates within it that most warrant further examination.

Proton pump inhibitors

PPI outcomes divided cleanly by organ system. Gastrointestinal symptoms were female-predominant: nausea (ROR 1.84, 95% CI 1.80-1.89), abdominal pain (1.33, 1.28-1.38) and diarrhoea (1.32, 1.29-1.35). Renal outcomes were male-predominant: acute kidney injury (0.80, 0.77-0.82), tubulointerstitial nephritis (0.83, 0.78-0.89) and chronic kidney disease (0.95, 0.93-0.98), as was hypomagnesaemia (0.79, 0.74-0.84). Hyponatraemia showed no sex difference (1.04, 0.99-1.09).

All four male-predominant renal and electrolyte estimates lie above the severity baseline of 0.63, so they are not explained by it.

Alopecia across unrelated classes

The two largest estimates in this study, and in the programme as a whole, were for alopecia: 8.34 in TNF blockers and 8.15 in GLP-1 receptor agonists. These classes share no mechanism, indication or patient population, and their confidence intervals overlap almost entirely. Alopecia is a recognised adverse effect of both, but an eight-fold female excess in reporting, reproduced to within 2% across two unrelated classes, is unlikely to represent an eight-fold difference in occurrence.

Discussion

This study was designed to test the limits of a reporting baseline rather than to confirm it, and it found them. In two of three new classes the death reporting odds ratio fell within the range established across seven previous classes. In the third it did not, and the departure was large: 0.38 against a nine-class mean of 0.632, more than seven standard deviations away.

The failure is informative rather than merely inconvenient. GLP-1 receptor agonists are also the outlier on absolute seriousness, with barely a quarter of female reports flagged serious against roughly half to four-fifths in every other class [2, 5]. A substantial and growing share of use is for weight reduction in people who are not otherwise ill, and much reporting appears to come from that population [1, 4]. The baseline therefore describes a reporting population rather than the database as a whole, and it can be expected to shift wherever that population shifts. For practice, this means the baseline should be estimated within the class under study rather than assumed from published values, and we would revise our earlier framing accordingly.

The alopecia finding is the most striking individual result. An eight-fold female excess, reproduced to within two percent across two classes with nothing pharmacological in common, is difficult to read as a difference in incidence [3]. Hair loss is both more visible and more consequential for many women, and is more likely to be noticed, attributed to a medicine and reported [10]. This is the clearest instance in the programme of spontaneous reporting measuring the salience of an event rather than its occurrence, and it has a practical corollary: disproportionality signals for cosmetically salient outcomes should be interpreted with particular caution in sex-stratified analyses.

The PPI results extend a pattern now observed in four classes. Women disproportionately reported gastrointestinal symptoms while men disproportionately reported renal injury, reproducing the symptom-versus-structural divergence previously seen in muscle with statins, in the gastrointestinal tract with NSAIDs, and in infection severity with TNF blockers here [5]. That this structure recurs across unrelated classes and organ systems argues for a general mechanism in how symptomatic and structural events reach reporting systems, rather than four coincidences.

Two class-specific observations merit follow-up in data with a denominator. The male predominance of acute kidney injury with GLP-1 agonists (0.48) sits well below even the lowered baseline for that class, and the male predominance of renal and electrolyte disturbance with PPIs is consistent across four distinct terms.

Study limitations: Spontaneous reporting yields no denominator of exposed patients, so no incidence or risk can be calculated, and a reporting odds ratio measures disproportionate reporting rather than relative risk. Causality is not established for individual reports. We could not adjust for age, indication, dose, duration or concomitant medication. This is a particular limitation for TNF blockers, where the underlying indications differ by sex, and for GLP-1 agonists, where the balance between diabetes and weight-reduction use almost certainly differs by sex and may itself explain part of the class's anomalous profile. We can show that the GLP-1 reporting population is distinctive but cannot characterise it directly. Three TNF blocker outcomes of clinical importance, including tuberculosis and lymphoma, could not be estimated because the female count fell below the retrievable range while the male count did not. Their direction is known and their magnitude is not; reporting a bound would have been misleading, and they are omitted. The baseline remains characterised rather than decomposed. Establishing whether it arises from reporter channel, healthcare contact, severity threshold or underlying event distribution requires simultaneous stratification by sex, class and reporter qualification, which the public interface cannot perform because it rejects queries combining three field restrictions. Replication using the downloadable quarterly FAERS files would permit it, and the GLP-1 anomaly makes that the highest-priority next step in this programme.

Conclusion

Across 1,753,642 reports for three classes chosen for their dissimilarity, male reports were again more often serious, making ten classes of ten. The death reporting baseline held for TNF blockers and proton pump inhibitors but failed decisively for GLP-1 receptor agonists, a class whose reporting population is also the least serious of any examined. The baseline is therefore a property of a reporting population rather than a constant, and should be estimated within each study rather than assumed. An eight-fold female excess in reported alopecia, reproduced almost exactly in two unrelated classes, illustrates how strongly the salience of an event can shape what a pharmacovigilance database contains.

Author Contributions: Conceptualization, N.G.; methodology, N.G.; software, N.G.; validation, K.T.; formal analysis, N.G.; investigation, N.G.; resources, K.T.; data curation, K.T.; writing—original draft preparation, S.Z N.G.; writing—review and editing, K.T.; visualization, N.G.; supervision, N.G.; project administration, N.G. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data are publicly available through the openFDA Drug Adverse Event API (https://open.fda.gov/apis/drug/event/). More data related to this research could be obtained upon a reasonable request to the corresponding author.

Acknowledgments: The authors would like to acknowledge Allama Iqbal Medical College, Lahore, for providing research facilities.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rushovich, T., Gompers, A., et al. Adverse drug events by sex after adjusting for baseline rates of drug use. JAMA Network Open. **2023**. 6(8): e2329074 10.1001/jamanetworkopen.2023.29074.
2. Food, U.S. and Drug, A. openFDA: Drug Adverse Event API. **2025**.
3. Rothman, K.J., Lanes, S., and Sacks, S.T. The reporting odds ratio and its advantages over the proportional reporting ratio. Pharmacoepidemiology and Drug Safety. **2004**. 13(8): 519–523 10.1002/pds.1001.
4. Sakaeda, T., Tamon, A., et al. Data mining of the public version of the FDA Adverse Event Reporting System. International Journal of Medical Sciences. **2013**. 10(7): 796–803 10.7150/ijms.6048.
5. Drici, M.D. and Clement, N. Is gender a risk factor for adverse drug reactions? The example of drug-induced long QT syndrome. Drug Safety. **2001**. 24(8): 575–585 10.2165/00002018-200124080-00002.
6. Evans, S.J.W., Waller, P.C., and Davis, S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiology and Drug Safety. **2001**. 10(6): 483–486 10.1002/pds.677.
7. van Puijenbroek, E.P., Bate, A., et al. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Pharmacoepidemiology and Drug Safety. **2002**. 11(1): 3–10 10.1002/pds.668.
8. Benjamini, Y. and Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Statistical Methodology). **1995**. 57(1): 289–300 10.1111/j.2517-6161.1995.tb02031.x.
9. von Elm, E., Altman, D.G., et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. The Lancet. **2007**. 370(9596): 1453–1457 10.1016/S0140-6736(07)61602-X.
10. Montastruc, J.L., Sommet, A., et al. Benefits and strengths of the disproportionality analysis for identification of adverse drug reactions in a pharmacovigilance database. British Journal of Clinical Pharmacology. **2011**. 72(6): 905–908 10.1111/j.1365-2125.2011.04037.x.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of RES Publishers and/or the editor(s). RES Publishers and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.