Understanding the consequences of GP referral thresholds: taking the instrumental approach

Jen Lewis ,[1] Christopher Burton [2]

[1]School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK
[2]Academic Unit of Primary Medical Care, University of Sheffield, Sheffield, UK

Correspondence to
Professor Christopher Burton, Academic Unit of Primary Medical Care, University of Sheffield, Sheffield, UK; chris.burton@sheffield.ac.uk

Linked

► http://dx.doi.org/10.1136/bmjqs-2022-014944

Check for updates

Trade-offs between patient safety and efficient use of healthcare services occur in clinical decisions across all forms of healthcare. In the case of acutely unwell older patients, decisions about referral to hospital involve trade-offs between the safety associated with inpatient hospital treatment and the burden on both the pa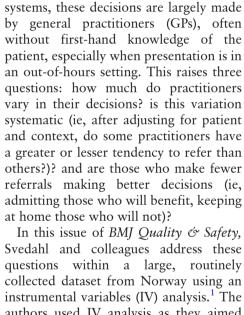tient and health system associated with hospital admission. In many healthcare systems, these decisions are largely made by general practitioners (GPs), often without first-hand knowledge of the patient, especially when presentation is in an out-of-hours setting. This raises three questions: how much do practitioners vary in their decisions? is this variation systematic (ie, after adjusting for patient and context, do some practitioners have a greater or lesser tendency to refer than others?)? and are those who make fewer referrals making better decisions (ie, admitting those who will benefit, keeping at home those who will not)?

In this issue of *BMJ Quality & Safety,* Svedahl and colleagues address these questions within a large, routinely collected dataset from Norway using an instrumental variables (IV) analysis.[1] The authors used IV analysis as they aimed to delineate the *causal* relationship, rather than simply to show an association, between referral by out-of-hours GPs and older patients' subsequent use of healthcare services and mortality up to 6 months. This is important because the relationship between referral and mortality depends on both the patients' initial condition (sicker patients are more likely to be admitted, thus introducing confounding by indication) and the treatment (following referral) they receive.

Nearly 500 000 patients aged 65 years or older were included in the study. While all patients were included, the nature of the analysis (explained further) facilitates a focus on those patients whose referrals could be attributed to their GP's 'referral threshold' or tendency to refer more or fewer patients. For these referred patients, the authors found that there was increased subsequent use of health services including outpatient specialist clinics and primary care physicians, and reduced mortality up to 6 months. This was taken to imply that while lower physician referral thresholds (tendency to refer *more* patients) would lead to increased subsequent service use, they also result in lower short-term and medium-term mortality, and consequently, that thresholds should not be raised without a clear assessment of the accuracy of referral decisions.

USING IV ANALYSIS TO EVALUATE RELATIONSHIPS
To understand these findings and place them in context, it is imperative to understand the advantages and potential pitfalls of this type of analysis. IV analysis is able to support inferences regarding the casual effect of one or more explanatory variables on an outcome by using a third variable—that is associated with the explanatory variable but not directly with the outcome—as the 'instrument' variable in the analysis. It helps to account for both measured *and* unmeasured confounding variables, making it an attractive option in a situation where a randomised controlled trial is either unfeasible or unethical, and where it is not possible to measure or include all possible confounders in an analysis. With greater availability and use of large routine datasets, IV analysis is an increasingly popular approach in health research.[2]

However, the key issue is whether the instrument variable used is in fact a good instrument. There are several important assumptions that the data must meet for IV analysis to be a valid approach, including

that the instrument variable must affect the explanatory variable of interest, must not affect the outcome except via its effects on the explanatory variable and that the instrument is not otherwise associated with the outcome via other covariates (either measured or unmeasured).[3] If IV analysis does not meet these assumptions or is otherwise used inappropriately, spurious conclusions can and do result.[4] It is complex to design and to implement, and assumptions may not be easy to meet or to evidence.[5][6] A suitable instrument may be difficult to identify or may not exist at all. To be useful, an instrument must impact on as many levels of the explanatory variable as possible. Furthermore, inclusion of an instrument does not always remove the need to adjust for covariates, nor is it inevitably an improvement over a standard covariate-adjusted regression.[7]

Returning to our three questions about variation in referrals by GPs, the study by Svedahl and colleagues used an IV to describe the referral threshold or tendency of each practitioner to admit older adults. This threshold was calculated as the proportion of older adults not known to the practitioner who were referred during out-of-hours work. The authors found that there was variation in this threshold between practitioners and that it appeared to be independent of patient factors such as age or comorbidities that would determine the need for referral; that is, it was systematic—attributable to the practitioner rather than the patients they saw.

Our third question was whether low-referring clinicians were making better decisions (ie, selecting patients most likely to benefit from referral and so reducing overall costs without impacting outcomes for patients who were not referred). The authors compared the IV approach with a conventional multivariable analysis. The multivariable approach suggested that patients who were referred had higher short-term and medium-term risk of death (Hazard Ratio (HR) 1.41 for days 0–10, HR 1.55 for days 0–180), which may be unsurprising if the sickest people are the ones being referred. In contrast, the IV analysis revealed a lower risk of death after referral for patients whose GP had greater referral tendency, compared with patients treated by low-referring doctors who had worse survival (HR 0.53 for days 0–10, HR 0.72 for days 0–180). While a difference in short-term mortality could have arisen from low-referring doctors making appropriate shared decisions about managing patients near the end of life at home rather than in the hospital (where more intervention might prolong life by only a month or two), the observed survival difference at 180 days suggests that lower referral rates included some worse decisions—not referring some patients who might have gained sustained benefit.

Both analyses were adjusted for important patient and visit characteristics to improve the strength of causal inferences. Importantly, the study did not just include the most vulnerable patients: it included all patients over 65, excluded those with multiple out-of-hours contacts in recent months as they will be more likely to be known to physicians, and had only 10% of patients with more than one significant comorbidity recorded.

IMPLICATIONS FOR EVALUATING PATIENT SAFETY IN COMPLEX ROUTINE DATA

The markedly different findings between the IV analysis and multivariable regression analysis remind us of the importance of an appropriate analysis and demonstrate how striking the effect can be for our conclusions, particularly when interrogating observational data and in situations where a research question or analysis plan was absent at the stage of data collection. The potential utility of IV analysis is obvious, but care must be taken with its use and interpretation.

In this case, the authors make an extremely thorough attempt to discover possible violations of assumptions, perform sensitivity analyses to examine their choice of instrument and take care to explicate that the conclusions apply only to those patients whose referral is due to the GP referral threshold. However, even in this convincing case, we must still be cautious. One thing we do not have detail on is the overall quality of the data in terms of cleanliness and missingness, which is important to judge the subsequent quality of any analysis.[8] Additionally, instrument–outcome confounders beyond those investigated here—with the potential to derail such an analysis—are plausible and do exist, for example, patient health behaviours,[9] reminding us that an exhaustive verification of assumptions is at best difficult, may not be possible with certain datasets and yet may have a profound impact on the analysis. It is also not clear how large the group of patients is to whom the conclusions relate, and thus how broadly applicable the instrument is; that is, some patients are unwell enough that all GPs will refer, some well enough that none will, so what proportion of patients fall into the category with unclear referral indications, for whom GP referral threshold is relevant?

Additionally, as with any relatively novel method, results should be compared with an appropriate comparator—whether this has been done here is questionable. The results shown by Svedahl and colleagues are striking because they are in the opposite direction to the comparator analysis (the multivariable regression), but it is not clear in this case that this comparator has been adequately adjusted using appropriate variables. In the context of the present research question, it is surprising that no measure of general health or comorbidity is included in this comparator. The absence of such important confounders in the multivariable regression analysis results in referral being associated with greater risk of mortality when in reality, the opposite is true.[10] In standard practice, such variables would almost certainly be included in such an analysis, meaning that the current conclusion drawn from the multivariate analysis would be extremely unlikely.

The authors state it may not be possible to account for a sufficient set of covariates to indicate the patient case mix. That is fair, but it would still seem astute to include any available important covariates. This is probably, therefore, not an entirely realistic comparison of analyses, may overstate the relative contribution of IV analysis and underestimate the utility of a simpler analysis that adequately adjusts for the most relevant covariates.

CLINICAL IMPLICATIONS OF INDIVIDUAL REFERRAL THRESHOLDS

Attempting to understand variation in clinical decisions around referral has a long history[11] and remains a live problem.[12] Variation in general practice referral of patients through suspected cancer pathways can be better explained by tendency to refer (or 'referral thresholds') than by variation in diagnostic accuracy for referrals.[13] Variation in referral tendency or threshold is plausible, but how should we characterise it? The absence of any relationship between tendency to refer and a wide range of patient characteristics suggests that it is not simply bias against a particular group of patients. Rather it may reflect what Kahneman *et al* has recently characterised as one of the several forms of 'noise' in human judgement.[14] If that is the case, only in-depth analysis of individuals' decision making is likely to clarify its nature. Whatever the cause, the scale of the impact implied by Svedahl *et al*'s analysis[1] suggests this is important. Based on the analysis reported here, GPs working for out-of-hours services might benefit from knowing how their referral rate compared with their peers (and nudging it up if it was on the low side). Furthermore, any guidance about more conservative referral policies should be viewed with caution and implemented and evaluated in ways which permit the early detection of signals indicating harm.

CONCLUSIONS

Inevitably, there are further questions. Would more complex mixtures of covariates reduce the GP referral tendency effect? How big a difference does this have in real terms? What would it take to produce a meaningful shift in behaviour or outcomes? Are there differences between the patients who are subsequently admitted to the hospital compared with those who are seen as acute outpatients?

These questions are unlikely to be answerable through randomised controlled trials, and so methods of analysis that allow the most unbiased examination of observational data possible are required. This study makes an important contribution in this area, but we must remember that no amount of comprehensive analysis can make up for a lack of quality data. Collecting routine data in a thorough, well-structured and research-amenable manner should be a priority to help delineate the consequences of referral in complex cases and further support evidence-based policy.

ORCID iDs
Jen Lewis http://orcid.org/0000-0002-3765-1566
Christopher Burton http://orcid.org/0000-0003-0233-2431

REFERENCES

1 Svedahl ER, Pape K, Austad B, *et al*. Impact of altering referral threshold from out-of-hours primary care to hospital on patient safety and further health service use: a cohort study. *BMJ Qual Saf* 2023;32:330–40.

2 Widding-Havneraas T, Chaulagain A, Lyhmann I, *et al*. Preference-based instrumental variables in health research rely on important and underreported assumptions: a systematic review. *J Clin Epidemiol* 2021;139:269–78.

3 Lousdal ML. An introduction to instrumental variable assumptions, validation and estimation. *Emerg Themes Epidemiol* 2018;15:1.

4 Soumerai SB, Koppel R. The reliability of instrumental variables in health care effectiveness research: less is more. *Health Serv Res* 2017;52:9–15.

5 Widding-Havneraas T, Zachrisson HD. A gentle introduction to instrumental variables. *J Clin Epidemiol* 2022;149:203–5.

6 Pokropek A. Introduction to instrumental variables and their application to large-scale assessment data. *Large-Scale Assess Educ* 2016;4:1–20.

7 Ceyisakar IE, van Leeuwen N, Steyerberg EW, *et al*. Instrumental variable analysis to estimate treatment effects: a simulation study showing potential benefits of conditioning on hospital. *BMC Med Res Methodol* 2022;22:121.

8 Kilkenny MF, Robinson KM. Data quality: "garbage in - garbage out." *Health Inf Manag* 2018;47:103–5.

9 Garabedian LF, Chu P, Toh S, *et al*. Potential bias of instrumental variable analyses for observational comparative effectiveness research. *Ann Intern Med* 2014;161:131–8.

10 Julious SA, Mullee MA. Confounding and simpson's paradox. *BMJ* 1994;309:1480–1.

11 Moore AT, Roland MO. How much variation in referral rates among general practitioners is due to chance? *BMJ* 1989;298:500–2.

12 Abel G, Elliott MN. Identifying and quantifying variation between healthcare organisations and geographical regions: using mixed-effects models. *BMJ Qual Saf* 2019;28:1032–8.

13 Burton CD, McLernon DJ, Lee AJ, *et al*. Distinguishing variation in referral accuracy from referral threshold: analysis of a national dataset of referrals for suspected cancer. *BMJ Open* 2017;7:e016439.

14 Kahneman D, Sibony O, Sunstein CR. *Noise: a flaw in human judgment*. 1st ed. London: William Collins, 2021: ix–454.